

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 14:52:18 ; Search time 43 Seconds
(without alignments)
45.406 Million cell updates/sec

Title: US-09-807-665-41
Perfect score: 33
Sequence: 1 KSADLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues
Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	33	100.0	7	13 US-10-080-100-1	Sequence 1, Appli
2	33	100.0	7	13 US-10-080-100-4	Sequence 4, Appli
3	33	100.0	35	14 US-10-422-934-22	Sequence 22, Appl
4	33	100.0	67	10 US-09-500-700-39	Sequence 39, Appl
5	33	100.0	98	10 US-09-500-700-42	Sequence 42, Appl
6	33	100.0	105	10 US-09-765-555-74	Sequence 74, Appl
7	33	100.0	189	9 US-09-908-153B-34	Sequence 34, Appl
8	33	100.0	414	14 US-10-157-899A-2	Sequence 2, Appli
9	33	100.0	414	14 US-10-157-899A-6	Sequence 6, Appli
10	33	100.0	414	14 US-10-157-899A-8	Sequence 8, Appli
11	33	100.0	414	14 US-10-157-899A-10	Sequence 10, Appl
12	33	100.0	438	14 US-10-157-899A-4	Sequence 4, Appli
13	33	100.0	438	14 US-10-157-899A-12	Sequence 12, Appl
14	33	100.0	438	14 US-10-157-899A-14	Sequence 14, Appl
15	33	100.0	438	14 US-10-157-899A-16	Sequence 16, Appl

Sequence 2043, Ap
Sequence 2045, Ap
Sequence 442, App
Sequence 23, Appl
Sequence 46, Appl
Sequence 55496, A
Sequence 56471, A
Sequence 58058, A
Sequence 59772, A
Sequence 67445, A
Sequence 73172, A
Sequence 75636, A
Sequence 20, Appl
Sequence 68855, A
Sequence 77981, A
Sequence 156346,
Sequence 262520,
Sequence 95, Appl
Sequence 303, App
Sequence 303, App
Sequence 56563, A
Sequence 6, Appli
Sequence 6, Appli
Sequence 40, Appl
Sequence 8, Appli
Sequence 332, App
Sequence 634, App
Sequence 634, App
Sequence 220, App
Sequence 2431, Ap

ALIGNMENTS

RESULT 1
US-10-080-100-1
; Sequence 1, Application US/10080100
; Publication No. US20020165356A1
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos
; APPLICANT: Dreier, Birgit
; TITLE OF INVENTION: Zinc Finger Binding Domains for
; TITLE OF INVENTION: Nucleotide Sequence ANN
; FILE REFERENCE: TSRI 760.0
; CURRENT APPLICATION NUMBER: US/10/080,100
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US/09/791,106
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthesized
US-10-080-100-1

Query Match 100.0%; Score 33; DB 13; Length 7;
Best Local Similarity 100.0%; Pred.No. 1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 1 KSADLKR 7

RESULT 2
US-10-080-100-4
; Sequence 4, Application US/10080100
; Publication No. US20020165356A1
; GENERAL INFORMATION:

```
; APPLICANT: Barbas, Carlos
; APPLICANT: Dreier, Birgit
; TITLE OF INVENTION: Zinc Finger Binding Domains for
; TITLE OF INVENTION: Nucleotide Sequence ANN
; FILE REFERENCE: TSRI 760.0
; CURRENT APPLICATION NUMBER: US/10/080,100
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US/09/791,106
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthesized
US-10-080-100-4

Query Match      100.0%; Score 33; DB 13; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
      |||||
Db      1 KSADLKR 7

RESULT 3
US-10-422-934-22
; Sequence 22, Application US/10422934
; Publication No. US20030186841A1
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos F., III
; APPLICANT: Kadan, Michael
; APPLICANT: Beerli, Roger
; TITLE OF INVENTION: LIGAND ACTIVATED TRANSCRIPTIONAL REGULATOR PROTEINS
; FILE REFERENCE: 22908-1227C
; CURRENT APPLICATION NUMBER: US/10/422,934
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: 09/586,625
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 09/433,042
; PRIOR FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
US-10-422-934-22

Query Match      100.0%; Score 33; DB 14; Length 35;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
      |||||
Db      20 KSADLKR 26

RESULT 4
US-09-500-700-39
; Sequence 39, Application US/09500700
; Publication No. US20030059767A1
; GENERAL INFORMATION:
; APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
; APPLICANT: BARBAS III, Carlos F.
; APPLICANT: GOTTESFELD, Joel M.
; APPLICANT: WRIGHT, Peter E.
```

```
; TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES AND METHODS THEREFOR
; FILE REFERENCE: SCRIP1160-4
; CURRENT APPLICATION NUMBER: US/09/500,700
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: US 08/863,813
; PRIOR FILING DATE: 1997-05-27
; PRIOR APPLICATION NUMBER: US 08/676,318
; PRIOR FILING DATE: 1996-12-30
; PRIOR APPLICATION NUMBER: PCT/US95/00829
; PRIOR FILING DATE: 1995-01-18
; PRIOR APPLICATION NUMBER: US 08/312,604
; PRIOR FILING DATE: 1994-09-28
; PRIOR APPLICATION NUMBER: US 08/183,119
; PRIOR FILING DATE: 1994-01-18
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Construction of C7 zinc finger
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (67)..(67)
; OTHER INFORMATION: Xaa is Tyr-Ala-Cys-Pro-Val-Glu-Ser-Cys-Asp-Arg-Arg-Phe-Ser-Lys-
; OTHER INFORMATION: -Ser-Ala-Asp-Leu-Lys-Arg-His-Ile-Arg-Ile-His-Thr-Gly-Glu-Lys-Pro
; OTHER INFORMATION: could be repeated 10 times
US-09-500-700-39

Query Match      100.0%; Score 33; DB 10; Length 67;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
      |||||
Db      20 KSADLKR 26
```

Cont'd. Page 3

```
RESULT 5
US-09-500-700-42
; Sequence 42, Application US/09500700
; Publication No. US20030059767A1
; GENERAL INFORMATION:
; APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
; APPLICANT: BARBAS III, Carlos F.
; APPLICANT: GOTTESFELD, Joel M.
; APPLICANT: WRIGHT, Peter E.
; TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES AND METHODS THEREFOR
; FILE REFERENCE: SCRIP1160-4
; CURRENT APPLICATION NUMBER: US/09/500,700
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: US 08/863,813
; PRIOR FILING DATE: 1997-05-27
; PRIOR APPLICATION NUMBER: US 08/676,318
; PRIOR FILING DATE: 1996-12-30
; PRIOR APPLICATION NUMBER: PCT/US95/00829
; PRIOR FILING DATE: 1995-01-18
; PRIOR APPLICATION NUMBER: US 08/312,604
; PRIOR FILING DATE: 1994-09-28
; PRIOR APPLICATION NUMBER: US 08/183,119
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-500-700-42

Query Match      100.0%; Score 33; DB 10; Length 98;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 KSADLKR 7
| | | | |
Db 20 KSADLKR 26

RESULT 6

US-09-765-555-74
; Sequence 74, Application US/09765555
; Publication No. US20030037355A1
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; TITLE OF INVENTION: Methods and compositions to modulate
; TITLE OF INVENTION: expression in plants
; FILE REFERENCE: 27801-20014.40
; CURRENT APPLICATION NUMBER: US/09/765,555
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 09/620,897
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 60/177,468
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of 3 finger protein C7

Query Match 100.0%; Score 33; DB 10; Length 105;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
| | | | |
Db 22 KSADLKR 28

RESULT 7

US-09-908-153B-34
; Sequence 34, Application US/09908153B
; Patent No. US20020168714A1
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos F.
; APPLICANT: Beeril, Roger
; APPLICANT: Schopfer, Ulrich
; TITLE OF INVENTION: REGULATION OF GENE EXPRESSION USING
; TITLE OF INVENTION: SINGLE-CHAIN, MONOMERIC, LIGAND DEPENDENT POLYPEPTIDE
; TITLE OF INVENTION: SWITCHES
; FILE REFERENCE: TSRI 725.1
; CURRENT APPLICATION NUMBER: US/09/908,153B
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 09/619,063
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; FEATURE: Synthesized
; NAME/KEY: VARIANT
; LOCATION: 1, 189
; OTHER INFORMATION: Xaa = Any Amino Acid

Query Match 100.0%; Score 33; DB 9; Length 189;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
| | | | |
Db 22 KSADLKR 28

RESULT 8

US-10-157-899A-2
; Sequence 2, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of the human-Estrogen-receptor-ligand-binding-domain and
; OTHER INFORMATION: zinc finger array(C7)

Query Match 100.0%; Score 33; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
| | | | |
Db 22 KSADLKR 28

RESULT 9

US-10-157-899A-6
; Sequence 6, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)

US-10-157-899A-6

```
Query Match      100.0%; Score 33; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
      |||||
Db      22 KSADLKR 28

RESULT 10
US-10-157-899A-8
; Sequence 8, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
US-10-157-899A-8

Query Match      100.0%; Score 33; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
      |||||
Db      22 KSADLKR 28

RESULT 11
US-10-157-899A-10
; Sequence 10, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
```

```
; OTHER INFORMATION: and a zinc finger array(C7)
US-10-157-899A-10

Query Match      100.0%; Score 33; DB 14; Length 414;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
      |||||
Db      22 KSADLKR 28

RESULT 12
US-10-157-899A-4
; Sequence 4, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of the human-Estrogen-receptor-ligand-binding-domain and
; OTHER INFORMATION: zinc finger array(C7)
US-10-157-899A-4

Query Match      100.0%; Score 33; DB 14; Length 438;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
      |||||
Db      22 KSADLKR 28

RESULT 13
US-10-157-899A-12
; Sequence 12, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; TITLE OF INVENTION: LIGANDS AND PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 438
; TYPE: PRT
```

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
US-10-157-899A-12
Query Match      100.0%; Score 33; DB 14; Length 438;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      22 KSADLKR 28

RESULT 14
US-10-157-899A-14
; Sequence 14, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
US-10-157-899A-14
Query Match      100.0%; Score 33; DB 14; Length 438;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      22 KSADLKR 28

RESULT 15
US-10-157-899A-16
; Sequence 16, Application US/10157899A
; Publication No. US20030143559A1
; GENERAL INFORMATION:
; APPLICANT: Bracken, Kathryn Rene
; APPLICANT: de los Angeles, Joseph Ernest
; APPLICANT: Huang, Ying
; APPLICANT: Kadan, Michael Joseph
; APPLICANT: Ksander, Gary Michael
; APPLICANT: Zerby, Dennis
; TITLE OF INVENTION: NOVEL ESTROGEN RECEPTOR LIGAND BINDING DOMAIN VARIANTS AND NOVEL
; FILE REFERENCE: 4-32018A
; CURRENT APPLICATION NUMBER: US/10/157,899A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/294,839
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
```

```

; SEQ ID NO 16
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion of a mutated human-Estrogen-receptor-ligand-binding-domain
; OTHER INFORMATION: and a zinc finger array(C7)
US-10-157-899A-16
Query Match      100.0%; Score 33; DB 14; Length 438;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      22 KSADLKR 28

Search completed: May 24, 2004, 14:58:02
Job time : 43 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 14:41:07 ; Search time 55 Seconds
(without alignments)
35.961 Million cell updates/sec

Title: US-09-807-665-41
Perfect score: 33
Sequence: 1 KSADLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	7	3 AAB02900	Aab02900 Nucleotid
2	33	100.0	7	5 ABP53161	Abp53161 Mouse tra
3	33	100.0	7	5 ABP53158	Abp53158 Mouse tra
4	33	100.0	7	6 ABU60707	Abu60707 Zinc fing
5	33	100.0	7	6 ABU60704	Abu60704 Zinc fing
6	33	100.0	35	4 AAE02027	Aae02027 Murine Zi
7	33	100.0	64	2 AAR83491	Aar83491 Two copie
8	33	100.0	64	2 AAW84390	Aaw84390 Zig268 C7
9	33	100.0	98	2 AAR83489	Aar83489 C7 zinc f
10	33	100.0	98	2 AAW84393	Aaw84393 The three
11	33	100.0	105	4 AAE06026	Aae06026 Three fin
12	33	100.0	109	2 AAW84455	Aaw84455 Sequence
13	33	100.0	189	5 AAB47900	Aab47900 Zinc fing
14	33	100.0	189	6 ABU66293	Abu66293 Zinc fing
15	33	100.0	354	2 AAR88802	Aar88802 11 copie
16	33	100.0	414	6 AAE35277	Aae35277 C7LBDAS f
17	33	100.0	414	6 AAE35279	Aae35279 C7LBDAS f
18	33	100.0	414	6 AAE35278	Aae35278 C7LBDAS f
19	33	100.0	414	6 AAE35275	Aae35275 C7LBDAS f
20	33	100.0	438	6 AAE35282	Aae35282 C7LBDAS f
21	33	100.0	438	6 AAE35281	Aae35281 C7LBDAS f
22	33	100.0	438	6 AAE35280	Aae35280 C7LBDAS f
23	33	100.0	438	6 AAE35276	Aae35276 C7LBDAS f
24	30	90.9	543	4 AAB20313	Aab20313 Soybean a
25	29	87.9	7	3 AAB02903	Aab02903 Nucleotid

26	29	87.9	47	3	AAB16844	Aab16844 Bacteriop
27	28	84.8	29	4	AAE02028	Aae02028 Murine Zi
28	28	84.8	155	4	AAB88501	Aab88501 Haemophil
29	28	84.8	155	5	AAU91432	Aau91432 Haemophil
30	28	84.8	155	6	ABU28547	Abu28547 Protein e
31	28	84.8	155	6	ABU31848	Abu31848 Protein e
32	28	84.8	155	6	ABU30134	Abu30134 Protein e
33	28	84.8	155	6	ABU39521	Abu39521 Protein e
34	28	84.8	155	6	ABU45248	Abu45248 Protein e
35	28	84.8	155	6	ABU27572	Abu27572 Protein e
36	28	84.8	155	6	ABU47712	Abu47712 Protein e
37	28	84.8	156	6	ABU40931	Abu40931 Protein e
38	28	84.8	156	6	ABU50057	Abu50057 Protein e
39	28	84.8	157	6	ABM69955	Abm69955 Photorhab
40	28	84.8	203	6	ABM72549	Abm72549 Staphyloc
41	28	84.8	204	4	AAG82682	Aag82682 S. epider
42	28	84.8	206	5	ABP39585	Abp39585 Staphyloc
43	28	84.8	395	3	AAG27000	Aag27000 Arabidops
44	28	84.8	421	3	AAG26999	Aag26999 Arabidops
45	28	84.8	440	3	AAG26998	Aag26998 Arabidops

ALIGNMENTS

RESULT 1
AAB02900
ID AAB02900 standard; peptide; 7 AA.
XX AAB02900;
AC AAB02900;
XX 18-SEP-2000 (first entry)
DT Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:41.
DE Zinc finger domain; alpha helix; nucleotide binding; DNA binding;
XX polydactyl protein; asymmetric target recognition;
KW gene specific transcriptional regulator; gene activator; gene repressor;
KW transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;
KW transgenic animal; antiviral; anticancer; diagnosis.
XX Synthetic.
OS WO200023464-A2.
PN 27-APR-2000.
PD 14-OCT-1999; 99WO-EP007742.
PF 16-OCT-1998; 98US-00173941.
PR (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PA (SCRI) SCRIPPS RES INST.
XX Barbas CF;
PI WPI; 2000-339648/29.
XX Novel isolated and purified zinc finger nucleotide-binding proteins with
PT specificity for GNN triplet sequences, useful in gene therapy and for
PT regulating gene function.
XX Disclosure; Fig 1; 48pp; English.
XX The invention relates to zinc finger nucleotide-binding proteins which
CC comprise 2-12, preferably 2-6, operatively linked motifs selected from
CC sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha
CC helical regions of zinc finger domains which specifically bind to target
CC nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked
CC by the peptide linker TGEKP (AAB02970). The Cys2-His2 zinc finger motif
CC is the most frequently utilised nucleic acid binding motif in eukaryotes,
CC and constitutes a beta-beta-alpha fold. Nucleic acid recognition is

CC achieved through specific contacts from side chains of amino acid
CC residues in the alpha helix. Each zinc finger can recognise a subsite of
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains
CC allows the recognition of extended contiguous asymmetric DNA sequences.
CC For example, a synthetic polydactyl protein containing six zinc finger
CC domains can recognise an 18 bp sequence, and such proteins are
CC potentially highly gene-specific. The novel nucleotide-binding zinc
CC finger proteins may therefore be used in the development of artificial
CC gene-specific transcriptional regulators. Such transcriptional switches
CC may be used to regulate the expression of oncogenes such as erbB-2,
CC overexpression of which is involved in malignant transformation. The
CC proteins are therefore useful in the treatment of cancers, and may also
CC be used to activate genes involved in fighting diseases, and to treat
CC viral infections by inhibiting the synthesis of viral gene products. They
CC may be used in DNA-based diagnostic applications. The proteins may also
CC be used in producing functional gene knockout or activation in
CC heterozygous transgenic animals. Proteins of the invention can
CC discriminate between sequences which have a single base difference. This
CC is manifested in a >100-fold decrease in affinity for the variant
CC sequence. Gene activation and repression can be achieved by targeting
CC within the gene transcript, suggesting that information obtained from
CC expressed sequence tags may be sufficient for the construction of gene
CC switches. Sequences AAB02876-302869 represent zinc finger alpha helix
CC phage library peptides disclosed in the invention
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 33; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 1 KSADLKR 7

RESULT 2
ABP53161
ID ABP53161 standard; peptide; 7 AA.

AC ABP53161;

DT 12-NOV-2002 (first entry)

DE Mouse transcription factor Zif268 zinc finger helix peptide SEQ ID NO:4.

XX Zinc finger binding domain; zinc finger nucleotide binding peptide;
KW zinc finger; gene expression; modulation; promoter; viral; Lentivirus;
KW human T-cell lymphocytic virus; HTLV; human immunodeficiency virus; HIV.

OS Mus sp.

XX WO200266640-A2.

XX 29-AUG-2002.

PF 21-FEB-2002; 2002WO-EP001862.

XX 21-FEB-2001; 2001US-00791106

XX (NOVS) NOVARTIS AG.

PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PA (SCRI) SCRIPPS RES INST.

PI Barbas CF, Dreier B;

XX WPI; 2002-674941/72.

XX New polypeptides comprising 2 to 10 zinc finger-nucleotide binding
PT peptides, useful for modulating gene expression in promoters from viral
PT groups, e.g. Lentivirus group, such HTLV-1 and 2, or HIV-1 and 2.

PS Example 1; Fig 1; 48pp; English.

XX The present invention describes a polypeptide comprising 2 to 12 zinc
CC finger-nucleotide binding peptides in which at least one contains a
CC nucleotide binding region. Also described: (1) an isolated and purified
CC polynucleotide encoding the polypeptide cited above; (2) an expression
CC vector containing the polynucleotide in (1); and (3) a process of
CC regulating expression of a nucleotide sequence that contains the sequence
CC S1 comprising exposing the nucleotide sequence to the polypeptide cited
CC above. (S1) is (5'-ANN)n-3', where n = any integer from 2-12; and N = A,
CC C, G, or T. The polypeptide is useful for modulating gene expression in
CC promoters from viral groups, e.g. Lentivirus group, such as human T-cell
CC lymphocytic virus (HTLV) 1 and 2, or human immunodeficiency virus (HIV) 1
CC or 2. ABQ75687 to ABQ75697 and ABP53112 to ABP53221 represent sequences
CC given in the exemplification of the present invention
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 33; DB 5; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 1 KSADLKR 7

RESULT 3
ABP53158
ID ABP53158 standard; peptide; 7 AA.

AC ABP53158;

DT 12-NOV-2002 (first entry)

DE Mouse transcription factor Zif268 zinc finger helix peptide SEQ ID NO:1.

XX Zinc finger binding domain; zinc finger nucleotide binding peptide;
KW zinc finger; gene expression; modulation; promoter; viral; Lentivirus;
KW human T-cell lymphocytic virus; HTLV; human immunodeficiency virus; HIV.

OS Mus sp.

XX WO200266640-A2.

XX 29-AUG-2002.

PF 21-FEB-2002; 2002WO-EP001862.

XX 21-FEB-2001; 2001US-00791106.

XX (NOVS) NOVARTIS AG.

PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PA (SCRI) SCRIPPS RES INST.

PI Barbas CF, Dreier B;

XX WPI; 2002-674941/72.

XX New polypeptides comprising 2 to 10 zinc finger-nucleotide binding
PT peptides, useful for modulating gene expression in promoters from viral
PT groups, e.g. Lentivirus group, such HTLV-1 and 2, or HIV-1 and 2.

PS Example 1; Fig 1; 48pp; English.

XX The present invention describes a polypeptide comprising 2 to 12 zinc
CC finger-nucleotide binding peptides in which at least one contains a
CC nucleotide binding region. Also described: (1) an isolated and purified
CC polynucleotide encoding the polypeptide cited above; (2) an expression
CC vector containing the polynucleotide in (1); and (3) a process of
CC regulating expression of a nucleotide sequence that contains the sequence
CC S1 comprising exposing the nucleotide sequence to the polypeptide cited
CC above. (S1) is (5'-ANN)n-3', where n = any integer from 2-12; and N = A,
CC C, G, or T. The polypeptide is useful for modulating gene expression in

CC promoters from viral groups, e.g. Lentivirus group, such as human T-cell
CC lymphocytic virus (HTLV) 1 and 2, or human immunodeficiency virus (HIV) 1
CC or 2. ABQ75687 to ABQ75697 and ABP53112 to ABP53221 represent sequences
CC given in the exemplification of the present invention
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 33; DB 5; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 1 KSADLKR 7

RESULT 4
ABU60707
ID ABU60707 standard; peptide; 7 AA.

XX AC ABU60707;
XX DT 06-MAY-2003 (first entry)
XX DE Zinc finger motif from mouse Zif268/C7 #4.
XX KW Zinc finger; mouse; Zif268/C7; phage display; ant-HIV; virucide;
XX KW HIV infection; human T cell lymphotropic virus infection; HTLV.
XX OS Mus sp.
XX PN US2002165356-A1.
XX PD 07-NOV-2002.
XX PF 21-FEB-2002; 2002US-00080100.
XX PR 21-FEB-2001; 2001US-00367356.
XX PA (SCRI) SCRIPPS RES INST.
XX PI Barbas CF, Dreier B;
XX DR WPI; 2003-255225/25.

XX PT New zinc finger-nucleotide binding polypeptides, useful in the treatment
XX of disorders associated with the aberrant expression or activity of the
XX polypeptide, such as HIV infection or human T cell lymphotropic virus
XX infection.
XX PS Example 1; Fig 1; 22pp; English.

XX CC The invention relates a new polypeptide comprising from 2-12 zinc finger-
XX nucleotide binding peptides at least one of which contains a nucleotide
XX binding region having any of 71 nucleotide binding sequences, fully
XX defined in the specification. Also included are an isolated and purified
XX polynucleotide that encodes the polypeptide, an expression vector
XX containing the polynucleotide and a process of regulating expression of a
XX nucleotide sequence that contains the sequence (5'-ANN)n-3', where n is
XX an integer from 2-12, the process comprising exposing the nucleotide
XX compositions of the present invention, are useful for modulating zinc-
XX finger protein binding and hence regulating aspects of gene expression.
XX The zinc finger polypeptides are useful in the treatment of treatment of
XX disorders associated with the aberrant expression or activity of the
XX polypeptide, such as viral infection, in particular HIV or human T cell
XX lymphotropic virus (HTLV) infection. The present sequence is a
XX nucleotide binding sequence (zinc finger) from the mouse transcription
XX factor Zif268/C7. These peptides were used to model zinc finger binding
XX and to design a phage display library of synthetic zinc fingers which may
XX be assembled into the zinc finger protein of the invention

SQ Sequence 7 AA;

Query Match 100.0%; Score 33; DB 6; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Query Match 100.0%; Score 33; DB 6; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 1 KSADLKR 7
RESULT 5
ABU60704
ID ABU60704 standard; peptide; 7 AA.
XX AC ABU60704;
XX DT 06-MAY-2003 (first entry)
XX DE Zinc finger motif from mouse Zif268/C7 #1.
XX KW Zinc finger; mouse; Zif268/C7; phage display; ant-HIV; virucide;
XX KW HIV infection; human T cell lymphotropic virus infection; HTLV.
XX OS Mus sp.
XX PN US2002165356-A1.
XX PD 07-NOV-2002.
XX PF 21-FEB-2002; 2002US-00080100.
XX PR 21-FEB-2001; 2001US-00367356.
XX PA (SCRI) SCRIPPS RES INST.
XX PI Barbas CF, Dreier B;
XX DR WPI; 2003-255225/25.

XX PT New zinc finger-nucleotide binding polypeptides, useful in the treatment
XX of disorders associated with the aberrant expression or activity of the
XX polypeptide, such as HIV infection or human T cell lymphotropic virus
XX infection.
XX PS Example 1; Fig 1; 22pp; English.

XX CC The invention relates a new polypeptide comprising from 2-12 zinc finger-
XX nucleotide binding peptides at least one of which contains a nucleotide
XX binding region having any of 71 nucleotide binding sequences, fully
XX defined in the specification. Also included are an isolated and purified
XX polynucleotide that encodes the polypeptide, an expression vector
XX containing the polynucleotide and a process of regulating expression of a
XX nucleotide sequence that contains the sequence (5'-ANN)n-3', where n is
XX an integer from 2-12, the process comprising exposing the nucleotide
XX compositions of the present invention, are useful for modulating zinc-
XX finger protein binding and hence regulating aspects of gene expression.
XX The zinc finger polypeptides are useful in the treatment of treatment of
XX disorders associated with the aberrant expression or activity of the
XX polypeptide, such as viral infection, in particular HIV or human T cell
XX lymphotropic virus (HTLV) infection. The present sequence is a
XX nucleotide binding sequence (zinc finger) from the mouse transcription
XX factor Zif268/C7. These peptides were used to model zinc finger binding
XX and to design a phage display library of synthetic zinc fingers which may
XX be assembled into the zinc finger protein of the invention

SQ Sequence 7 AA;

Db 1 KSADLKR 7
|||||
RESULT 6
AAE02027
ID AAE02027 standard; peptide; 35 AA.
XX AC
XX AC
DT 31-JUL-2001 (first entry)
XX DE
DE Murine Zif-C7 ZFP modified fragment #1.
XX
KW Fusion protein; nucleotide-binding domain; NBD; ligand-binding domain;
KW LBD; transcription regulating domain; TRD; zinc finger protein; ZFP;
KW ligand-activated transcriptional regulator; gene regulation;
KW gene therapy; cell proliferative disorder; cancer; psoriasis;
KW pemphigus vulgaris; Behcet's syndrome; lipid histiocytosis; Zif-C7;
KW murine.
XX
OS Mus sp.
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT Region 31..35
FT /label= Linker
XX
XX WO200130843-A1.
XX
PD 03-MAY-2001.
XX
XX 23-OCT-2000; 2000WO-EP010430.
XX
PR 25-OCT-1999; 99US-00433042.
PR 02-JUN-2000; 2000US-00586625.
XX
XX (NOVS) NOVARTIS AG.
PA (SCRI) SCRIPPS RES INST.
XX
XX Barbas CF, Kadan M, Beerli R;
PI WPI; 2001-308618/32.
DR
XX New fusion protein containing nucleotide-binding and ligand-binding
PT domains, useful e.g. in gene therapy of cancer, provides ligand-activated
PT control of gene expression.
XX
PS Disclosure; Page 49; 218pp; English.
XX
CC The invention relates to fusion protein comprising a nucleotide-binding
CC domain (NBD), a ligand-binding domain (LBD) of an intracellular receptor
CC (ICR) and a transcription regulating domain (TRD). NBD is a polydactyl
CC zinc finger protein (ZFP), or a modular part of it, that interacts
CC specifically with a contiguous sequence of at least 3 nucleotides. The
CC fusion protein functions as a ligand-activated transcriptional regulator.
CC The fusion protein and the nucleic acid encoding it, are used to regulate
CC gene expression, particularly in gene therapy for treating malignant cell
CC proliferative diseases (e.g. colon cancer, prostate cancer, renal-cell
CC carcinoma) and non-malignant cell proliferative diseases (e.g. psoriasis,
CC pemphigus vulgaris, Behcet's syndrome and lipid histiocytosis). The
CC fusion protein and its DNA are also useful for treating diseases caused
CC by viruses in humans/plants, genetic and/or acquired diseases. The fusion
CC protein can be designed to target any selected gene (endogenous or
CC exogenous), and can be made to have different selectivity or specificity
CC for endogenous or exogenous ligands. The present sequence is murine Zif-
CC C7 ZFP fragment comprising modified finger 1 which has altered DNA
CC binding specificity
XX
SQ Sequence 35 AA;

Query Match 100.0%; Score 33; DB 4; Length 35;
Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSADLKR 7
Db 20 KSADLKR 26
RESULT 7
AAR83491
ID AAR83491 standard; protein; 64 AA.
XX AC
XX AAR83491;
XX 18-APR-1996 (first entry)
DT
XX Two copies of zinc finger variant C7 fused via consensus linker.
DE
XX zinc finger; variant; fusion protein; consensus linker;
KW multifinger protein; cancer; therapy; virus infection;
KW transcription-activator; transcription-inhibitor; plant disease.
XX
XX Synthetic.
OS
OS
FH Key Location/Qualifiers
FT Domain 7..30
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 31..35
FT /label= consensus_linker
FT Domain 36..59
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 60..64
FT /label= consensus_linker
FT /note= "sequence of terminal linker is subject to change
FT as it is not involved in linking two fingers together"
XX
PN WO9519431-A1.
XX
XX 20-JUL-1995.
XX
XX 18-JAN-1995; 95WO-US000829.
XX
PR 18-JAN-1994; 94US-00183119.
PR 28-SEP-1994; 94US-00312604.
XX
PA (SCRI) SCRIPPS RES INST.
XX
XX Barbas CF, Gottesfeld JM, Wright PE;
PI WPI; 1995-263862/34.
XX
DR New zinc finger-nucleotide binding polypeptide(s) - used for modulating
XX the function of cellular nucleotide sequences, partic. for treating cell
PT proliferative disorders.
PT
XX Example 13; Page 86; 135pp; English.
PS
XX Mutagenesis of the Zif268 protein produced a variant (designated C7) with
CC an enhanced affinity for the GCG-GCG-GCG target sequence. Proteins
CC containing 2 to 12 copies of the C7 finger, linked via the consensus
CC linker TGEKP, were constructed and were shown to have specificity for
CC their predicted targets. The present sequence is that of the protein
CC containing 2 copies of C7. Fingers utilised in the multifinger proteins
CC need not be identical and may be mixed to produce proteins which
CC recognise a desired target sequence. They may also be used with leucine
CC zippers to produce proteins with extended sequence recognition. The
CC proteins are useful for suppressing transcription and therefore for
CC treating proliferative disorders
XX
SQ Sequence 64 AA;

Query Match 100.0%; Score 33; DB 2; Length 64;

Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 20 KSADLKR 26

RESULT 8

AAW84390

ID AAW84390 standard; protein; 64 AA.
XX
AC AAW84390;
XX
DT 22-MAR-1999 (first entry)
XX
DE Zif268 C7 finger sequence.
XX
KW Zif268 three finger protein; zinc finger; nucleotide-binding protein;
KW cell proliferative disorder; gene therapy; cancer; psoriasis;
KW pemphigus vulgaris; Bechet's syndrome; lipid histiocytosis;
KW human immune deficiency virus; HIV; viral infection; transgenic plant.
XX

OS Synthetic.

PN WO9854311-A1.

XX 03-DEC-1998.

XX 27-MAY-1998; 98WO-US010801.

XX 27-MAY-1997; 97US-00863813.

XX (SCRI) SCRIPPS RES INST.

XX Barbas CF, Gottesfeld JM, Wright PE;

XX WPI; 1999-059831/05.

XX New zinc finger nucleotide-binding protein variant that modulates
PT selected nucleotide sequence - used for treatment of proliferative and
PT viral diseases by gene therapy, and can be made selective for any target
PT sequence.

XX Example 13; Page 89; 158pp; English.

XX The present sequence represents the C7 finger of the zif268 zinc three
CC finger nucleotide-binding protein. The specification describes zinc
CC finger nucleotide-binding protein variants with at least two zinc finger
CC modules that bind to a cellular nucleotide sequence and modulate its
CC function. Zinc finger proteins, and compositions containing them, are
CC used to increase or reduce transcription of a gene linked to the cellular
CC nucleotide sequence. The proteins are used specifically for treating or
CC preventing cell proliferative disorders (in humans, animals or plants,
CC including those induced by viruses), particularly where expressed from
CC nucleic acid by gene therapy (including ex vivo methods). Typical
CC diseases that can be treated are many forms of cancer, psoriasis,
CC pemphigus vulgaris, Bechet's syndrome and lipid histiocytosis, also
CC treatment of human immune deficiency virus (HIV) and other viral
CC infections, and production of transgenic plants resistant to bacterial
CC and viral diseases

XX Sequence 64 AA;

Query Match 100.0%; Score 33; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 20 KSADLKR 26

RESULT 9
AAW83489

ID AAR83489 standard; protein; 98 AA.
XX

AC AAR83489;

XX 23-MAR-1996 (first entry)

XX C7 zinc finger 3-finger fusion protein.

XX C7 mutant; zinc finger-1; 3-finger fusion protein; mouse; Zif268;
KW immediate-early protein; Krox-24; HIV-1; AIDS; cancer; therapy;
KW transcription-activator; transcription-inhibitor; virus infection;
KW plant disease.

XX Synthetic.

PN WO9519431-A1.

XX 20-JUL-1995.

XX 18-JAN-1995; 95WO-US000829.

XX 18-JAN-1994; 94US-00183119.

XX 28-SEP-1994; 94US-00312604.

XX (SCRI) SCRIPPS RES INST.

XX Barbas CF, Gottesfeld JM, Wright PE;

XX WPI; 1995-263862/34.

XX N-PSDB; AAT01842.

XX New zinc finger-nucleotide binding polypeptide(s) - used for modulating
PT the function of cellular nucleotide sequences, partic. for treating cell
PT proliferative disorders.

XX Example 13; Fig 15; 135pp; English.

XX The sequence represents a fusion protein with 3 copies of zinc finger
CC protein C7, a variant of mouse immediate-early protein Zif268 (Krox-24),
CC in which the zinc finger-1 specificity and affinity are modified. Clone
CC C7 is a mutant which is improved 13-fold in affinity for binding to the
CC wild-type sequence GCG, and specificity of the C7 protein is improved 9-
CC fold for an HIV-1 target sequence. The 3-finger fusion protein binds to
CC its target sequence in the oligonucleotide hairpin AAT01855 with improved
CC affinity (9 nM). The fusion protein may be used for specific activation
CC or repression of transcription, and may be used in therapy of a cell
CC proliferative disorder, e.g. cancer or virus infection, or to treat plant
CC disease

XX Sequence 98 AA;

Query Match 100.0%; Score 33; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 20 KSADLKR 26

RESULT 10

AAW84393

ID AAW84393 standard; protein; 98 AA.

XX

AC AAW84393;

XX 22-MAR-1999 (first entry)

XX The three zinc finger C7 protein.

XX zinc finger; nucleotide-binding protein; cell proliferative disorder;

KW gene therapy; cancer; psoriasis; pemphigus vulgaris; Bechet's syndrome;
KW lipid histiocytosis; human immune deficiency virus; HIV; viral infection;
KW transgenic plant; C7.

XX Synthetic.

XX Key Location/Qualifiers
FH Misc-difference 91
FT /note= "encoded by CAC"
FT Misc-difference 98
FT /note= "encoded by ACT"

XX WO9854311-A1.

XX 03-DEC-1998.

XX 27-MAY-1998; 98WO-US010801.

XX 27-MAY-1997; 97US-00863813.

XX (SCRI) SCRIPPS RES INST.

XX Barbas CF, Gottesfeld JM, Wright PE;

XX WPI; 1999-059831/05.

DR N-PSDB; AAX03253.

XX New zinc finger nucleotide-binding protein variant that modulates

PT selected nucleotide sequence - used for treatment of proliferative and
PT viral diseases by gene therapy, and can be made selective for any target
PT sequence.

XX Example 13; Fig 15; 158pp; English.

XX The present sequence represents a three zinc finger C7 protein, made in
CC the course of the invention. The specification describes zinc finger
CC nucleotide-binding protein variants with at least two zinc finger modules
CC that bind to a cellular nucleotide sequence and modulate its function.
CC Zinc finger proteins, and compositions containing them, are used to
CC increase or reduce transcription of a gene linked to the cellular
CC nucleotide sequence. The proteins are used specifically for treating or
CC preventing cell proliferative disorders (in humans, animals or plants,
CC including those induced by viruses), particularly where expressed from
CC nucleic acid by gene therapy (including ex vivo methods). Typical
CC diseases that can be treated are many forms of cancer, psoriasis,
CC pemphigus vulgaris, Bechet's syndrome and lipid histiocytosis, also
CC treatment of human immune deficiency virus (HIV) and other viral
CC infections, and production of transgenic plants resistant to bacterial
CC and viral diseases

XX Sequence 98 AA;

Query Match 100.0%; Score 33; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 20 KSADLKR 26

RESULT 11

AAE06026

ID AAE06026 standard; protein; 105 AA.

XX AAE06026;

DT 24-SEP-2001 (first entry)

DE Three finger protein (C7).

XX Zinc finger protein; ZFP; gene expression; metabolic pathway regulator;
KW modulation; plant technology; agriculture; three finger protein; C7.

XX Unidentified.

XX WO200152620-A2.

XX 26-JUL-2001.

XX 19-JAN-2001; 2001WO-US001817.

XX 21-JAN-2000; 2000US-0177468P.

PR 21-JUL-2000; 2000US-00620897.

XX (SCRI) SCRIPPS RES INST.

PA (SYGN) SYNGENTA AGRIC DISCOVERY INC.

XX Barbas CF, Stege JT, Guan X, Dalmia B;

XX WPI; 2001-465325/50.

DR N-PSDB; AAD11616.

XX New zinc finger proteins, useful for modulating or regulating gene
PT expression and metabolic pathways in plants, e.g. for treating in the
PT plant cells a disorder that is associated with abnormal expression of the
PT target gene.

XX Disclosure; Page 137; 156pp; English.

XX The patent discloses methods and compositions to modulate the expression
CC of a target gene in plant cells. The method involves providing plant
CC cells with a zinc finger protein (ZFP) which is capable of specifically
CC binding to a target nucleotide sequence or its complementary strand
CC within a target gene and allowing the ZFP binding to the target
CC nucleotide sequence, where the expression of the target gene in the plant
CC cells is modulated. The ZFP and fusions of the ZFP proteins are useful
CC for modulating or regulating gene expression and metabolic pathways in
CC plants. The ZFP, fusion proteins and methods are useful in plant and
CC agricultural technology. The method is useful particularly for treating a
CC disorder in the plant cells, where the disorder is associated with
CC abnormal expression of the target gene. The present sequence is three
CC finger protein (C7)

XX Sequence 105 AA;

Query Match 100.0%; Score 33; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 22 KSADLKR 28

RESULT 12

AAW84455

ID AAW84455 standard; peptide; 109 AA.

XX AAW84455;

DT 22-MAR-1999 (first entry)

XX Sequence used to construct C7 zinc fingers.

XX zinc finger; nucleotide-binding protein; cell proliferative disorder;
KW gene therapy; cancer; psoriasis; pemphigus vulgaris; Bechet's syndrome;
KW lipid histiocytosis; human immune deficiency virus; HIV; viral infection;
KW transgenic plant; C7.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 76.109

FT /note= "this sequence is repeated 2-10 times"

XX

PN WO9854311-A1.
XX
PD 03-DEC-1998.
XX
PF 27-MAY-1998; 98WO-US010801.
XX
PR 27-MAY-1997; 97US-00863813.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Barbas CF, Gottesfeld JM, Wright PE;
XX
XX WPI; 1999-059831/05.
XX
XX New zinc finger nucleotide-binding protein variant that modulates
PT selected nucleotide sequence - used for treatment of proliferative and
PT viral diseases by gene therapy, and can be made selective for any target
PT sequence.
XX
PS Disclosure; Page 53; 158pp; English.
XX
CC The present sequence can be used to construct C7 zinc fingers, in the
CC course of the invention. The specification describes zinc finger
CC nucleotide-binding protein variants with at least two zinc finger modules
CC that bind to a cellular nucleotide sequence and modulate its function.
CC zinc finger proteins, and compositions containing them, are used to
CC increase or reduce transcription of a gene linked to the cellular
CC nucleotide sequence. The proteins are used specifically for treating or
CC preventing cell proliferative disorders (in humans, animals or plants,
CC including those induced by viruses), particularly where expressed from
CC nucleic acid by gene therapy (including ex vivo methods). Typical
CC diseases that can be treated are many forms of cancer, psoriasis,
CC pemphigus vulgaris, Bechet's syndrome and lipid histiocytosis, also
CC treatment of human immune deficiency virus (HIV) and other viral
CC infections, and production of transgenic plants resistant to bacterial
CC and viral diseases
XX
SQ Sequence 109 AA;

Query Match 100.0%; Score 33; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db |||||
20 KSADLKR 26

RESULT 13
AAB47900
ID AAB47900 standard; protein; 189 AA.
XX
AC AAB47900;
XX
DT 16-MAY-2002 (first entry)
XX
DE Zinc finger DNA binding domain, 2C7.
XX
KW Polymerase chain reaction; primer; amplify; PCR; reporter;
KW polypeptide gene switch; ligand binding domain; LBD; transcription;
KW nuclear hormone receptor; gene therapy; regulation; expression.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 1
FT /note= "Unspecified amino acid encoded by NNG"
FT Misc-difference 189
FT /note= "Unspecified amino acid encoded by CNN"
XX
PN WO200206463-A2.
XX
PD 24-JAN-2002.

XX 16-JUL-2001; 2001WO-EP008190.
XX
XX 18-JUL-2000; 2000US-00619063.
XX
XX (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PA (SCRI) SCRIPPS RES INST.
XX
XX Beerli R, Schopfer U, Barbas CF;
PI
XX WPI; 2002-179793/23.
DR N-PSDB; AAI72475.
XX
XX Single chain, monomeric polypeptide gene switches for regulating gene
PT function, comprise two ligand binding domains derived from nuclear
PT hormone receptors operatively linked to one functional domain.
XX
PS Disclosure; Fig 5; 63pp; English.
XX
CC This sequence represents the zinc finger DNA binding domain 2C7. This
CC protein could be used in the construction of a non-naturally occurring
CC polypeptide gene switch comprising two ligand binding domains (LBDs)
CC derived from nuclear hormone receptors operatively linked to a first
CC functional domain. The gene switch is useful for regulating the function
CC of a target nucleotide that contains a defined sequence, by exposing the
CC target nucleotide to the polypeptide gene switch in the presence of a
CC ligand that binds one of the LBDs of the polypeptide, where the DNA
CC binding domain of the polypeptide binds the defined sequence, or the
CC functional domain of the polypeptide alters the function of the target
CC nucleotide. The gene switch is also useful in the field of gene therapy
CC and as a regulator of gene expression or transcription. The advantage of
CC the gene switches of the invention over existing gene switches is the
CC need for only a single molecular switch and a single expression vector
CC for production of that switch
XX
SQ Sequence 189 AA;

Query Match 100.0%; Score 33; DB 5; Length 189;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db |||||
22 KSADLKR 28

RESULT 14
ABU66293
ID ABU66293 standard; protein; 189 AA.
XX
AC ABU66293;
XX
DT 23-MAY-2003 (first entry)
XX
DE Zinc finger binding domain 2C7.
XX
KW Zinc finger binding domain; gene switch; gene regulation.
XX
OS Unidentified.
XX
XX US2002168714-A1.
PN
XX
XX 14-NOV-2002.
XX
XX 18-JUL-2001; 2001US-00908153.
PF
XX 18-JUL-2000; 2000US-00325747.
PR
XX (SCRI) SCRIPPS RES INST.
PA
XX Barbas CF, Beerli R, Schopfer U;
PI
XX

DR WPI; 2003-328405/31.
DR N-PSDB; ACA02219.
XX
PT Novel polypeptide gene switch useful for regulating gene function,
PT comprises two ligand binding domains derived from nuclear hormone
PT receptors operatively linked to a functional domain.
XX
PS Disclosure; Fig 5; 33pp; English.
XX
CC The invention relates to a non-naturally occurring polypeptide (or
CC polypeptide gene switch) comprising two ligand binding domains derived
CC from nuclear hormone receptors operatively linked to a first functional
CC domain. The polypeptide is useful for regulating the function of a target
CC nucleotide that contains a defined sequence, by exposing the target
CC nucleotide to the polypeptide in the presence of a ligand that binds one
CC of the ligand binding domains of the polypeptide, where the DNA binding
CC domain of the polypeptide binds the defined sequence or alters the
CC function of the target nucleotide. The gene switches can be produced
CC using a single molecular switch and a single expression vector. The
CC present sequence represents the amino acid sequence of a zinc finger
CC binding domain
XX
SQ Sequence 189 AA;

Query Match 100.0%; Score 33; DB 6; Length 189;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 22 KSADLKR 28

RESULT 15
AAR88802
ID AAR88802 standard; protein; 354 AA.
XX AAR88802;
AC
XX 19-APR-1996 (first entry)
DT
XX 11 copies of zinc finger variant C7 fused via consensus linkers.
DE
XX zinc finger; variant; fusion protein; consensus linker;
KW multifinger protein; cancer; therapy; virus infection;
KW transcription-activator; transcription-inhibitor; plant disease.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Domain 7..30
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 31..35
FT /label= consensus_linker
FT Domain 36..59
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 60..64
FT /label= consensus_linker
FT Domain 65..88
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 89..93
FT /label= consensus_linker
FT Domain 94..117
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 118..122
FT /label= consensus_linker
FT Domain 123..146
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"

FT Region 147..151
FT /label= consensus_linker
FT Domain 152..175
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 176..180
FT /label= consensus_linker
FT Domain 181..204
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 205..209
FT /label= consensus_linker
FT Domain 210..233
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 234..238
FT /label= consensus_linker
FT Domain 239..262
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 263..267
FT /label= consensus_linker
FT Domain 268..291
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 292..296
FT /label= consensus_linker
FT Domain 297..320
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 321..325
FT /label= consensus_linker
FT Domain 326..349
FT /label= zinc_finger C7
FT /note= "binds GCG-GCG-GCG target sequence"
FT Region 350..354
FT /label= consensus linker
FT /note= "sequence of terminal linker is subject to change
FT as it is not involved in linking two fingers together"
XX
PN W09519431-A1.
XX 20-JUL-1995.
PD
XX 18-JAN-1995; 95WO-US000829.
PF
XX 18-JAN-1994; 94US-00183119.
PR 28-SEP-1994; 94US-00312604.
PR
XX (SCRI) SCRIPPS RES INST.
PA
XX Barbas CF, Gottesfeld JM, Wright PE;
PI WPI; 1995-263862/34.
XX
DR New zinc finger-nucleotide binding polypeptide(s) - used for modulating
XX the function of cellular nucleotide sequences, partic. for treating cell
XX proliferative disorders.
PT
XX Example 13; Page 86; 135pp; English.
PS
XX Mutagenesis of the Zif268 protein produced a variant (designated C7) with
CC an enhanced affinity for the GCG-GCG-GCG target sequence. Proteins
CC containing 2 to 12 copies of the C7 finger, linked via the consensus
CC linker TGEKP, were constructed and were shown to have specificity for
CC their predicted targets. The present sequence is that of the protein
CC containing 12 copies of C7. Fingers utilised in the multifinger proteins
CC need not be identical and may be mixed to produce proteins which
CC recognise a desired target sequence. They may also be used with leucine
CC zippers to produce proteins with extended sequence recognition. The
CC proteins are useful for suppressing transcription and therefore for
CC treating proliferative disorders
XX

SQ Sequence 354 AA;
Query Match 100.0%; Score 33; DB 2; Length 354;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSADLKR 7
Db 20 KSADLKR 26

Search completed: May 24, 2004, 14:50:58
Job time : 56 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 14:49:53 ; Search time 22 Seconds
(without alignments)
16.426 Million cell updates/sec

Title: US-09-807-665-41
Perfect score: 33
Sequence: 1 KSADLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	7	3	US-09-173-941-41
2	33	100.0	7	4	US-09-494-190-41
3	33	100.0	64	3	US-08-676-318A-39
4	33	100.0	76	3	US-08-863-813A-39
5	33	100.0	98	3	US-08-863-813A-42
6	33	100.0	98	3	US-08-676-318A-42
7	29	87.9	7	3	US-09-173-941-44
8	29	87.9	7	4	US-09-494-190-44
9	29	87.9	149	4	US-09-134-000C-6760
10	29	87.9	251	4	US-09-540-236-3483
11	28	84.8	73	3	US-08-485-355B-46
12	28	84.8	156	4	US-09-489-039A-12780
13	28	84.8	179	4	US-09-543-681A-5089
14	28	84.8	206	4	US-09-134-001C-4430
15	28	84.8	423	2	US-08-290-731C-11
16	28	84.8	652	1	US-08-318-831-8
17	28	84.8	1297	2	US-08-290-731C-4
18	28	84.8	1704	3	US-08-485-355B-40
19	28	84.8	1711	3	US-08-369-822C-10
20	28	84.8	1711	3	US-08-582-776C-10
21	28	84.8	1711	3	US-08-434-831B-10
22	27	81.8	180	4	US-09-543-681A-4912
23	27	81.8	198	4	US-09-252-991A-19295
24	27	81.8	661	4	US-09-371-338-7
25	26	78.8	7	4	US-09-989-789-711
26	26	78.8	68	2	US-08-460-890A-46
27	26	78.8	68	3	US-08-167-641C-46

28	26	78.8	68	3	US-08-460-971A-46	Sequence 46, Appl
29	26	78.8	68	3	US-08-462-040-46	Sequence 46, Appl
30	26	78.8	139	4	US-09-134-000C-4577	Sequence 4577, Ap
31	26	78.8	212	2	US-08-924-759-14	Sequence 14, Appl
32	26	78.8	212	3	US-09-248-335-14	Sequence 14, Appl
33	26	78.8	222	3	US-09-248-335-66	Sequence 66, Appl
34	26	78.8	226	4	US-09-533-029-88	Sequence 88, Appl
35	26	78.8	227	4	US-09-621-976-8	Sequence 8, Appl
36	26	78.8	230	4	US-09-107-532A-3897	Sequence 3897, Ap
37	26	78.8	295	4	US-09-134-001C-4242	Sequence 4242, Ap
38	26	78.8	343	1	US-08-463-090B-11	Sequence 11, Appl
39	26	78.8	576	4	US-09-543-681A-4990	Sequence 4990, Ap
40	26	78.8	828	4	US-09-107-532A-4191	Sequence 4191, Ap
41	26	78.8	894	4	US-09-543-681A-4724	Sequence 4724, Ap
42	26	78.8	1447	4	US-09-376-330-17	Sequence 17, Appl
43	26	78.8	1786	2	US-08-477-451-16	Sequence 16, Appl
44	25	75.8	7	4	US-09-989-789-1229	Sequence 1229, Ap
45	25	75.8	27	2	US-08-620-151-120	Sequence 120, App

ALIGNMENTS

RESULT 1
US-09-173-941-41
; Sequence 41, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV0081S
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-41

Query Match 100.0%; Score 33; DB 3; Length 7;
Best local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
|||
Db 1 KSADLKR 7

RESULT 2
US-09-494-190-41
; Sequence 41, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: phage display
OTHER INFORMATION: selected and mutagenized
US-09-494-190-41

Query Match 100.0%; Score 33; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
| | | | | | |
Db 1 KSADLKR 7

RESULT 3

US-08-676-318A-39
Sequence 39, Application US/08676318A
Patent No. 6242568
GENERAL INFORMATION:
APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES AND
METHODS THEREFOR
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,318A
FILING DATE: 30-DEC-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A., Ph.D.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 08401/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 64 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..64

US-08-676-318A-39

Query Match 100.0%; Score 33; DB 3; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
| | | | | | |
Db 20 KSADLKR 26

RESULT 4

US-08-863-813A-39
Sequence 39, Application US/08863813A
Patent No. 6140466
GENERAL INFORMATION:
APPLICANT: Barbas III, Carlos F.

APPLICANT: Gottesfeld, Joel M.
APPLICANT: Wright, Peter E.
TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES
TITLE OF INVENTION: AND METHODS THEREFOR
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/863,813A
FILING DATE: 27-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/676,318
FILING DATE: 18-JUL-1996
APPLICATION NUMBER: 08/183,119
FILING DATE: 18-JAN-1996
APPLICATION NUMBER: US95/00829
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A., Ph.D.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 08401/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
LOCATION: 76...76
OTHER INFORMATION: where Xaa at position 76 is Tyr-Ala-Cys-Pro-Val-Glu-Ser-Cys-Asp-Arg-Phe-Ser-Lys-Ser-Ala-Asp-Leu-Lys-Thr-Gly-Glu-Lys-Pro-Met-Lys-Leu-Glu-Pro-repeated 10 times
OTHER INFORMATION: amino acids; some Xaa's may be missing
US-08-863-813A-39

Query Match 100.0%; Score 33; DB 3; Length 76;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
| | | | | | |
Db 20 KSADLKR 26

RESULT 5

US-08-863-813A-42
Sequence 42, Application US/08863813A
Patent No. 6140466
GENERAL INFORMATION:
APPLICANT: Barbas III, Carlos F.
APPLICANT: Gottesfeld, Joel M.
APPLICANT: Wright, Peter E.
TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES
TITLE OF INVENTION: AND METHODS THEREFOR
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA

COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICANT: US/08/863,813A
FILING DATE: 27-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/676,318
FILING DATE: 18-JUL-1996
APPLICATION NUMBER: 08/183,119
FILING DATE: 18-JAN-1996
APPLICATION NUMBER: US95/00829
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A., Ph.D.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 08401/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 98 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-863-813A-42

Query Match 100.0%; Score 33; DB 3; Length 98;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 20 KSADLKR 26

RESULT 6
US-08-676-318A-42
Sequence 42, Application US/08676318A
Patent No. 6242568
GENERAL INFORMATION:
APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
TITLE OF INVENTION: ZINC FINGER PROTEIN DERIVATIVES AND
METHODS THEREFOR
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,318A
FILING DATE: 30-DEC-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A., Ph.D.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 08401/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070

TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 98 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-676-318A-42

Query Match 100.0%; Score 33; DB 3; Length 98;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 20 KSADLKR 26

RESULT 7
US-09-173-941-44
Sequence 44, Application US/09173941
Patent No. 6140081
GENERAL INFORMATION:
APPLICANT: BARBAS, Carlos F.
TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
FILE REFERENCE: NOV0081S
CURRENT APPLICATION NUMBER: US/09/173,941
CURRENT FILING DATE: 1998-10-16
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 44
LENGTH: 7
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: nucleotide
OTHER INFORMATION: codon binding sequence
US-09-173-941-44

Query Match 87.9%; Score 29; DB 3; Length 7;
Best Local Similarity 85.7%; Pred. No. 3e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 1 KSAELKR 7

RESULT 8
US-09-494-190-44
Sequence 44, Application US/09494190
Patent No. 6610512
GENERAL INFORMATION:
APPLICANT: BARBAS, Carlos F.
TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
FILE REFERENCE: TSRI 645.2
CURRENT APPLICATION NUMBER: US/09/494,190
CURRENT FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: EP/99/07742
PRIOR FILING DATE: 1999-10-14
PRIOR APPLICATION NUMBER: US 09/173,941
PRIOR FILING DATE: 1998-10-16
NUMBER OF SEQ ID NOS: 126
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 44
LENGTH: 7
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: phage display
OTHER INFORMATION: selected and mutagenized
US-09-494-190-44

```
Query Match      87.9%; Score 29; DB 4; Length 7;
Best Local Similarity 85.7%; Pred. No. 3e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      1 KSAELKR 7

RESULT 9
US-09-134-000C-6760
; Sequence 6760, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6760
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6760

Query Match      87.9%; Score 29; DB 4; Length 149;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      46 KXADLKR 52

RESULT 10
US-09-540-236-3483
; Sequence 3483, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3483
; LENGTH: 251
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-3483

Query Match      87.9%; Score 29; DB 4; Length 251;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      186 KDADLKR 192

RESULT 11
US-08-485-355B-46
; Sequence 46, Application US/08485355B
; Patent No. 6177075
; GENERAL INFORMATION:
; APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
; TITLE OF INVENTION: Insect Viruses and Their Uses in

; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,355B
; FILING DATE: 07-Jun-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/440,522
; FILING DATE: 12-MAY-1995
; APPLICATION NUMBER: US 08/089,372
; FILING DATE: 08-JUL-1993
; APPLICATION NUMBER: AU PL4081/92
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 73 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-08-485-355B-46

Query Match      84.8%; Score 28; DB 3; Length 73;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 SADLKR 7
Db      56 SADLKR 61

RESULT 12
US-09-489-039A-12780
; Sequence 12780, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12780
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12780

Query Match      84.8%; Score 28; DB 4; Length 156;
Best Local Similarity 71.4%; Pred. No. 60;
```

```
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSADLKR 7
Db 46 KNADIKR 52

RESULT 13
US-09-543-681A-5089
; Sequence 5089, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5089
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5089

Query Match 84.8%; Score 28; DB 4; Length 179;
Best Local Similarity 71.4%; Pred. No. 69;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSADLKR 7
Db 68 KNADIKR 74

RESULT 14
US-09-134-001C-4430
; Sequence 4430, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4430
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4430

Query Match 84.8%; Score 28; DB 4; Length 206;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KSADLKR 7
Db 96 KSKDLKR 102

RESULT 15
US-08-290-731C-11
; Sequence 11, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
```

```
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,
; TITLE OF INVENTION: AND MSOS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL0921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-731C-11

Query Match 84.8%; Score 28; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 SADLKR 7
Db 71 SADLKR 76

Search completed: May 24, 2004, 14:53:23
Job time : 23 secs
```

GenCore version 5.1.6
right (c) 1993 - 2004 Compugen Ltd.

OM protein - proteoin search, using sw model

Run on: May 24, 2004, 14:48:28 ; Search time 21 Seconds
(without alignments)
32.064 Million cell updates/sec

Title: US-09-807-665-41
Perfect score: 33
Sequence: 1 KSADLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	90.9	328	2 T37734	SUR4 family protein
2	30	90.9	560	2 T48841	heat shock protein
3	30	90.9	1626	2 A75613	hypothetical prote
4	29	87.9	648	2 T33339	hypothetical prote
5	29	87.9	1029	2 F87369	TonB-dependent rec
6	28	84.8	155	1 Q9ECF1	ybeA protein - Esc
7	28	84.8	155	2 G64140	hypothetical prote
8	28	84.8	155	2 B90713	hypothetical prote
9	28	84.8	155	2 F85563	hypothetical prote
10	28	84.8	155	2 AD0581	conserved hypothet
11	28	84.8	156	2 AI0317	conserved hypothet
12	28	84.8	180	2 B70202	cytidylate kinase
13	28	84.8	183	2 S26961	H+-transporting tw
14	28	84.8	321	2 T43455	hypothetical prote
15	28	84.8	348	1 S20911	alcohol dehydrogen
16	28	84.8	477	2 G71319	hypothetical prote
17	28	84.8	763	2 S35908	ATP-dependent Clp
18	28	84.8	803	2 T40475	bromodomain prote
19	28	84.8	818	2 T02436	proline-rich prote
20	28	84.8	856	2 T31234	traC protein homol
21	28	84.8	988	2 T18986	hypothetical prote
22	28	84.8	1217	2 F69823	probable phosphoes
23	28	84.8	1297	2 S25714	son-of-sevenless-2
24	28	84.8	1648	2 S57163	probable membrane
25	27	81.8	126	1 HSZPB2	histone H2B.1 - fi
26	27	81.8	174	2 S59848	ribosomal protein
27	27	81.8	180	2 S56056	ribosomal protein
28	27	81.8	195	2 T36854	hypothetical prote
29	27	81.8	208	2 D87565	conserved hypothet

30	27	81.8	249	2 A81892	probable transposa
31	27	81.8	331	2 AB1737	conserved hypothet
32	27	81.8	331	2 A11367	conserved hypothet
33	27	81.8	343	2 D71341	probable exported
34	27	81.8	353	2 F72394	glutamate 5-kinase
35	27	81.8	363	2 A86267	hypothetical prote
36	27	81.8	388	2 AG3256	cobW protein limpo
37	27	81.8	389	2 D90248	DNA topoisomerase
38	27	81.8	448	2 S71427	transcription fact
39	27	81.8	449	2 S41647	zinc finger 5 prot
40	27	81.8	513	2 A99568	atp synthase alpha
41	27	81.8	513	2 G81298	probable periplasm
42	27	81.8	520	2 A64033	hypothetical prote
43	27	81.8	561	2 H86442	unknown protein [i
44	27	81.8	552	2 H86221	hypothetical prote
45	27	81.8	666	2 T20375	hypothetical prote

ALIGNMENTS

RESULT 1

T37734
SUR4 family protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C:Accession: T37734; T39101
R:Wedler, H.; Dueterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21742
A:Accession: T37734
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-328 <WED>
A:Cross-references: EMBL:AL117213; PIDN:CAB55289.1; GSPDB:GN00066; SPDB:SPAC1639.01c
A:Experimental source: strain 972h-; cosmid cl639
R:Wedler, H.; Dueterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21827
A:Accession: T39101
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-328 <WE2>
A:Cross-references: EMBL:AL117212; PIDN:CAB55288.1; GSPDB:GN00066; SPDB:SPAC806.09c
A:Experimental source: strain 972h-; cosmid c806
C:Genetics:
A:Gene: SPAC806.09c; SPAC1639.01c
A:Map position: 1
A:Introns: 5/2; 125/3
C:Superfamily: Saccharomyces probable membrane protein YCR034w

Query Match 90.9%; Score 30; DB 2; Length 328;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7

Db 291 KNADLKR 297

RESULT 2

T48841
heat shock protein cct1 [similarity] - Haloferax volcanii
N:Alternate names: chaperonin-containing Tcp-1
C:Species: Haloferax volcanii
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 21-Jul-2000
C:Accession: T48841
R:Kuo, Y.P.; Thompson, D.K.; St Jean, A.; Charlebois, R.L.; Daniels, C.J.
J. Bacteriol. 179, 6318-6324, 1997
A:Title: Characterization of two heat shock genes from Haloferax volcanii: a model syst
A:Reference number: Z24371; MUID:97474254; PMID:9335278
A:Accession: T48841
A:Status: preliminary; translated from GB/EMBL/DDBJ

RESULT 7
G64140
YbeA protein homolog HI0033 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 12-Jun-2003
C:Accession: G64140
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: G64140
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-155 <TIGR>
A:Cross-references: GB:U32688; GB:L42023; NID:G1572966; PIDN:AAC21711.1; PID:G1572978; T
C:Genetics:
A:Start codon: GTG
C:Superfamily: uncharacterized conserved protein HI0033

Query Match 84.8%; Score 28; DB 2; Length 155;
Best Local Similarity 71.4%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
|:|:|
Db 45 KNADIKR 51

RESULT 8
B90713
Hypothetical protein ECs0674 [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050952)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 12-Jun-2003
C:Accession: B90713
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90713
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-155 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA034097.1; PID:G13360132; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: ECs0674
C:Superfamily: uncharacterized conserved protein HI0033

Query Match 84.8%; Score 28; DB 2; Length 155;
Best Local Similarity 71.4%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
|:|:|
Db 45 KNADIKR 51

RESULT 9
F85563
Hypothetical protein ybeA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 12-Jun-2003
C:Accession: F85563
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Faller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85563
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-155 <STO>
A:Cross-references: GB:AE005174; NID:G12513537; PIDN:AAG54970.1; GSPDB:GN00145; UWGP:Z0
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ybeA
C:Superfamily: uncharacterized conserved protein HI0033

Query Match 84.8%; Score 28; DB 2; Length 155;
Best Local Similarity 71.4%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
|:|:|
Db 45 KNADIKR 51

RESULT 10
AD0581
Conserved hypothetical protein STY0692 [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 12-Jun-2003
C:Accession: AD0581
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P. Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K. A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AD0581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-155 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05118.1; PID:G16501892; GSPDB:GN00176
C:Genetics:
A:Gene: STY0692
C:Superfamily: uncharacterized conserved protein HI0033

Query Match 84.8%; Score 28; DB 2; Length 155;
Best Local Similarity 71.4%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
|:|:|
Db 45 KNADIKR 51

RESULT 11
AI0317
Conserved hypothetical protein YPO2605 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 12-Jun-2003
C:Accession: AI0317
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AI0317
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92848.1; PID:G15980592; GSPDB:GN00175
C:Genetics:
A:Gene: YPO2605
C:Superfamily: uncharacterized conserved protein HI0033


```

Query Match      84.8%; Score 28; DB 2; Length 156;
Best Local Similarity 71.4%; Pred. No. 35;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      45 KNADIKR 51

RESULT 12
B70202
cytidylate kinase (cmk-2) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: B70202
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:980655943; PMID:9403685
A:Accession: B70202
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-180 <KLE>
A:Cross-references: GB:AE001180; GB:AE000783; NID:g2688755; PIDN:AAC67164.1; PID:g268875
A:Experimental source: strain B31
C:Superfamily: Methanococcus jannaschii cytidylate kinase

Query Match      84.8%; Score 28; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLK 6
Db      89 KSADLK 94

RESULT 13
S26961
H+-transporting two-sector ATPase (EC 3.6.3.14) delta chain - red alga (Antithamnion sp.
C:Species: chloroplast Antithamnion sp.
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 03-Jun-2002
C:Accession: S26961
R:Kostrzewa, M.; Zetsche, K.
J. Mol. Biol. 227, 961-970, 1992
A:Title: Large ATP synthase operon of the red alga Antithamnion sp. resembles the corres
A:Reference number: S26957; MUID:93021132; PMID:1404401
A:Accession: S26961
A:Molecule type: DNA
A:Residues: 1-183 <KOS>
A:Cross-references: EMBL:X63382; NID:g14170; PIDN:CAA44983.1; PID:g14176
C:Genetics:
A:Gene: atpD
A:Genome: Chloroplast
C:Superfamily: H+-transporting ATP synthase delta chain
C:Keywords: ATP biosynthesis; chloroplast; hydrolase; thylakoid

Query Match      84.8%; Score 28; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLK 6
Db      43 KSADLK 48

RESULT 14
T43455
hypothetical protein DKF2p34J1618.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000

```

```

C:Accession: T43455
R:Ottewaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z22515
A:Accession: T43455
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-321 <AAA>
A:Cross-references: EMBL:AL133637
A:Experimental source: adult testis; clone DKFZp34J1618
C:Genetics:
A:Note: DKFZp34J1618.1

Query Match      84.8%; Score 28; DB 2; Length 321;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 SADLKR 7
Db      74 SADLKR 79

RESULT 15
S20911
alcohol dehydrogenase (EC 1.1.1.1) II - yeast (Kluyveromyces marxianus var. lactis)
C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: S20911; S19804
R:Shain, D.H.; Salvatore, C.; Denis, C.L.
Mol. Gen. Genet. 232, 479-488, 1992
A:Title: Evolution of the alcohol dehydrogenase (ADH) genes in yeast: characterization
A:Reference number: S20911; MUID:92269769; PMID:1588917
A:Accession: S20911
A:Molecule type: DNA
A:Residues: 1-348 <SHA>
A:Cross-references: EMBL:X64397; NID:g2832; PIDN:CAA45739.1; PID:g2833
C:Genetics:
A:Gene: ADH2
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; NAD; oxidoreductase; zinc
F;29-336/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;173-202/Region: beta-alpha-beta NAD nucleotide-binding fold
F;44,67,154/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F;98,101,104,112/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match      84.8%; Score 28; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLK 6
Db      164 KSADLK 169

Search completed: May 24, 2004, 14:52:48
Job time : 22 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: May 24, 2004, 14:41:48 ; Search time 12 Seconds
(without alignments)
30.374 Million cell updates/sec

Title: US-09-807-665-41
Perfect score: 33
Sequence: 1 KSADLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	30	90.9	560 1	THS1_HALVO
2	28	84.8	155 1	Y033_HAEIN
3	28	84.8	155 1	Y033_ECOLI
4	28	84.8	155 1	YBEA_SALTY
5	28	84.8	155 1	YJ23_PASMU
6	28	84.8	156 1	YQ05_YERPE
7	28	84.8	180 1	KCY2_BORBU
8	28	84.8	183 1	ATPD_ANTSP
9	28	84.8	348 1	ADH2_KLULA
10	28	84.8	477 1	Y481_TREPA
11	28	84.8	763 1	CLPL_LACIA
12	28	84.8	970 1	K852_HUMAN
13	28	84.8	1217 1	YHCR_BACSU
14	28	84.8	1297 1	SOS2_MOUSE
15	28	84.8	1332 1	SOS2_HUMAN
16	28	84.8	1608 1	RRPL_BDV
17	28	84.8	1648 1	YJ9H_YEAST
18	27	81.8	125 1	H2B1_SCHPO
19	27	81.8	174 1	RL20_YEAST
20	27	81.8	297 1	YK46_METAC
21	27	81.8	343 1	TP38_TREPA
22	27	81.8	353 1	PROB_THEMEA
23	27	81.8	387 1	TP6A_SULTO
24	27	81.8	389 1	TP6A_SULSH
25	27	81.8	448 1	ZFG1_CHICK
26	27	81.8	449 1	ZFG1_HUMAN
27	27	81.8	449 1	ZFG1_MOUSE
28	27	81.8	482 1	EX1_BUCAP
29	27	81.8	513 1	ATPA_MYCPU
30	27	81.8	520 1	VG29_HAEIN
31	27	81.8	522 1	C5P3_ARATH
32	27	81.8	1191 1	NIFJ_RHORU
33	27	81.8	1616 1	YAED_SCHPO

34	26	78.8	148	1	CB20_XENLA	P52299 xenopus lae
35	26	78.8	161	1	HBPL_PARAD	P04396 parasponia
36	26	78.8	218	1	TRPF_SYNY3	P74435 synechocyst
37	26	78.8	253	1	VP8_FOWPV	P15913 fowlpox vir
38	26	78.8	309	1	KDGK_ECOLI	P37647 escherichia
39	26	78.8	325	1	HEMN_SYNVU	O31067 synechococc
40	26	78.8	338	1	GPDA_LACPL	Q88yk1 lactobacill
41	26	78.8	415	1	PROA_THEMEA	Q9wyc9 thermotoga
42	26	78.8	496	1	SRMS_MOUSE	Q62270 mus musculu
43	26	78.8	545	1	OXAA_NEIMA	Q9jw48 neisseria m
44	26	78.8	545	1	OXAA_NEIMB	Q9jxs4 neisseria m
45	26	78.8	549	1	TCPA_CABEL	P41988 caenorhabdi

ALIGNMENTS

RESULT 1	
THS1_HALVO	THS1_HALVO
ID	THS1_HALVO
AC	O30561; STANDARD; PRT; 560 AA.
DT	15-JUL-1998 (Rel. 36, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Thermosome subunit 1 (Heat shock protein CCT1).
GN	CCT1.
OS	Halobacterium volcanii (Haloferax volcanii).
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC	Halobacteriaceae; Haloferax.
OX	NCBI_TaxID=2246;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=97474254; PubMed=9335278;
RA	Kuo Y.-P., Thompson D.K., St Jean A., Charlebois R.L., Daniels C.J.;
RT	"Characterization of two heat shock genes from Haloferax volcanii: a
RT	model system for transcription regulation in the Archaea.";
RL	J. Bacteriol. 179:6318-6324(1997).
CC	-!- FUNCTION: MOLECULAR CHAPERONE; BINDS UNFOLDED POLYPEPTIDES IN
CC	VITRO, AND HAS A WEAK ATPASE ACTIVITY (BY SIMILARITY).
CC	-!- SUBUNIT: FORMS AN OLIGOMERIC COMPLEX OF EIGHT-MEMBERED RINGS
CC	(BY SIMILARITY).
CC	-!- INDUCTION: By heat shock.
CC	-!- SIMILARITY: Belongs to the TCP-1 chaperonin family.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL; AF010470; AAB81497.1; -.
DR	PIR; T48841; T48841.
DR	HSSP; P48424; 1A6D.
DR	InterPro; IPR002194; Chaperonin_TCP-1.
DR	InterPro; IPR001844; Chaperin_Cpn60.
DR	InterPro; IPR002423; Cpn60/TCP-1.
DR	InterPro; IPR008950; GroEL-ATPase.
DR	Pfam; PF00118; cpn60_TCP1; 1.
DR	PRINTS; PRO0298; CHAPERONIN60.
DR	PROSITE; PRO0304; TCOMPLEXTCP1.
DR	PROSITE; PS00750; TCP1_1; 1.
DR	PROSITE; PS00751; TCP1_2; 1.
DR	PROSITE; PS00995; TCP1_3; 1.
KW	Chaperone; ATP-binding; Heat shock.
SQ	SEQUENCE 560 AA; 58925 MW; 88B73B2AD70DC341 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 560;
Best Local Similarity 85.7%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7

```
Db          315 KSSDLKR 321
|||||
RESULT 2
Y033_HAEIN  STANDARD;          PRT;   155 AA.
AC P44470;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0247 protein HI0033.
GN HI0033.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: Belongs to the UPF0247 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC -----
DR EMBL; U32688; AAC21711.1; -.
DR PIR; G64140; G64140.
DR TIGR; HI0033; -.
DR HAMAP; MF 00658; -. 1.
DR InterPro; IPR003742; DUF163.
DR Pfam; PF02590; DUF163; 1.
DR TIGRFAMs; TIGR00246; TIGR00246; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 155 AA; 17250 MW; D147B9FA24389A1E CRC64;

Query Match      84.8%; Score 28; DB 1; Length 155;
Best Local Similarity 71.4%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY          1 KSDLKR 7
Db          45 KNADIKR 51
|||||
RESULT 3
YBEA_ECOLI  STANDARD;          PRT;   155 AA.
AC P05850;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0247 protein ybeA.
GN YBEA OR B0636 OR C0727 OR Z0782 OR ECS0674 OR SF0645 OR S0667.
OS Escherichia coli,
OS Escherichia coli O6,
```

```
OS Escherichia coli O157:H7, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 83334, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=87030266; PubMed=3533535;
RA Asoh S., Matsuzawa H., Ishino F., Strominger J.L., Matsuhashi M.,
RA Ohta T.;
RT "Nucleotide sequence of the bbpA gene and characteristics of the
RT deduced amino acid sequence of penicillin-binding protein 2 of
RT Escherichia coli K12.";
RL Eur. J. Biochem. 160:231-238(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli;
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M., Lin D.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [7]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
```

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [8]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
RT through comparison with genomes of *Escherichia coli* K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [9]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of *Shigella*
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (1.68 ANGSTROMS).
RA Benach J., Shen J., Rost B., Xiao R., Acton T., Montelione G.,
RA Hunt J.F.;
RT "Structure of ybeA from *E.coli*.";
RL Submitted (JAN-2003) to the PDB data bank.
CC -!- SIMILARITY: Belongs to the UPF0247 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04516; CAA28200.1; -;
CC EMBL; AE000168; AAC73737.1; -;
CC EMBL; U82598; AAB40836.1; -;
CC EMBL; D90704; BAA35283.1; -;
CC EMBL; AE016757; AAN79200.1; -;
CC EMBL; AE005242; AAG54970.1; -;
CC EMBL; AP002552; BAB34097.1; -;
CC EMBL; AE015095; AAN42281.1; -;
CC EMBL; AE016280; AAP16152.1; -;
CC PIR; B24995; QQECPI.
CC PIR; B90713; B90713.
CC PIR; F85563; F85563.
CC PDB; 1NS5; 04-FEB-03.
CC EcoGene; EGI1254; ybeA.
CC HAMAP; MF_00658; -; 1.
CC InterPro; IPR003742; DUF163.
CC Pfam; PF02590; DUF163; 1.
CC TIGRFAMs; TIGR00246; TIGR00246; 1.
CC Hypothetical protein; Complete proteome; 3D-structure.
CC CONFLICT 59 59 Q -> L (IN REF. 9).
CC SEQUENCE 155 AA; 17341 MW; DFA0EFCF4EB05E7 CRC64;
Query Match 84.8%; Score 28; DB 1; Length 155;
Best Local Similarity 71.4%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

|:|:|:

Db 45 KNADIKR 51
RESULT 4
YBEA_SALTY STANDARD; PRT; 155 AA.
AC Q8XF27;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, last sequence update)
DT 10-OCT-2003 (Rel. 42, last annotation update)
DE Hypothetical UPF0247 protein ybeA.
GN YBEA OR STM0641 OR STY0692 OR T2226.
OS *Salmonella typhimurium*, and
OS *Salmonella typhi*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Salmonella*.
OX NCBI_TaxID=502, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant *Salmonella*
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- SIMILARITY: Belongs to the UPF0247 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008725; AAL19592.1; -;
CC EMBL; AL627267; CAD05118.1; -;
CC EMBL; AE016841; AAO69829.1; -;
CC StyGene; SG?22??; ybeA.
CC HAMAP; MF_00658; -; 1.
CC InterPro; IPR003742; DUF163.
CC Pfam; PF02590; DUF163; 1.
CC TIGRFAMs; TIGR00246; TIGR00246; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 155 AA; 17399 MW; D34EFD3F4EB9D36A CRC64;
KW

```

Query Match      84.8%; Score 28; DB 1; Length 155;
Best Local Similarity 71.4%; Pred. No. 16;
Matches      5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      45 KNADIKR 51

RESULT 5
YJ23_PASMU
ID_YJ23_PASMU      STANDARD;      PRT;      155 AA.
AC_Q9CJR9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0247 protein PM1923.
GN PM1923.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.B., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- SIMILARITY: Belongs to the UPF0247 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF006229; AAK04007.1; -.
DR HAMAP; MF_00658; -; 1.
DR InterPro; IPR003742; DUF163.
DR Pfam; PF02590; DUF163; 1.
DR TIGRFAMs; TIGR00246; TIGR00246; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 155 AA; 17440 MW; 0A7E9B83D59435A7 CRC64;

Query Match      84.8%; Score 28; DB 1; Length 155;
Best Local Similarity 71.4%; Pred. No. 16;
Matches      5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      45 KNADIKR 51

RESULT 6
YQ05_YERPE
ID_YQ05_YERPE      STANDARD;      PRT;      156 AA.
AC_Q8ZDG3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0247 protein YPO2605/Y1179.
GN YPO2605 OR Y1179.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;

Query Match      84.8%; Score 28; DB 1; Length 155;
Best Local Similarity 71.4%; Pred. No. 16;
Matches      5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      45 KNADIKR 51

RESULT 7
KCY2_BORBU
ID_KCY2_BORBU      STANDARD;      PRT;      180 AA.
AC_O51759;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytidylate kinase 2 (EC 2.7.4.14) (CK 2) (Cytidine monophosphate
DE kinase 2) (CMP kinase 2).
GN CMK2 OR BB0819.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
```

```

RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- SIMILARITY: Belongs to the UPF0247 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AJ14153; CAC92848.1; -.
DR EMBL; AE013721; AAM84756.1; -.
DR PIR; A10317; A10317.
DR HAMAP; MF_00658; -; 1.
DR InterPro; IPR003742; DUF163.
DR Pfam; PF02590; DUF163; 1.
DR TIGRFAMs; TIGR00246; TIGR00246; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 156 AA; 17520 MW; 6132248883BFEEC2 CRC64;

Query Match      84.8%; Score 28; DB 1; Length 156;
Best Local Similarity 71.4%; Pred. No. 16;
Matches      5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 KSADLKR 7
Db      45 KNADIKR 51

RESULT 7
KCY2_BORBU
ID_KCY2_BORBU      STANDARD;      PRT;      180 AA.
AC_O51759;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytidylate kinase 2 (EC 2.7.4.14) (CK 2) (Cytidine monophosphate
DE kinase 2) (CMP kinase 2).
GN CMK2 OR BB0819.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
```

RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
CC -!- CATALYTIC ACTIVITY: ATP + (d)CMP = ADP + (d)CDP.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the cytidylate kinase family. Subfamily 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; AE001180; AAC67164.1; -.
DR PIR; B70202; B70202.
DR TIGR; BB0819; -.
DR HAMAP; MF_00239; -; 1.
KW Transferase; Kinase; ATP-binding; Complete proteome.
FT NP BIND 7 15 ATP (POTENTIAL).
SQ SEQUENCE 180 AA; 21066 MW; 4CD956B6F8A5900B CRC64;

Query Match 84.8%; Score 28; DB 1; Length 180;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLK 6
Db |||||
89 KSADLK 94

RESULT 8
ATPD_ANTSP STANDARD; PRT; 183 AA.
AC Q02849;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase delta chain, chloroplast (EC 3.6.3.14).
GN ATPD.
OS Antithamnion sp.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Ceramiaceae;
OC Antithamnion.
OX NCBI_TaxID=2767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LB 95.79;
RX MEDLINE=93021132; PubMed=1404401;
RA Kozrzewa M., Zetsche K.;
RT "Large ATP synthase operon of the red alga Antithamnion sp. resembles
RT the corresponding operon in cyanobacteria.";
RL J. Mol. Biol. 227:961-970(1992).
CC -!- FUNCTION: This protein seems to be part of the stalk that links
CC CF(0) to CF(1). It either transmits conformational changes from
CC CF(0) into CF(1) or is implicated in proton conduction.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC has three main subunits: a, b and c.
CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -!- SIMILARITY: Belongs to the ATPase delta chain family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; X63382; CAA44983.1; -.
DR PIR; S26961; S26961.
DR InterPro; IPR000711; ATPsynth_OSCP.
DR Pfam; PF00213; OSCP; 1.
DR PRINTS; PR00125; ATPASEDELTA.
DR TIGRfams; TIGR01145; ATP_synth_delta; 1.
DR PROSITE; PS00389; ATPASE_DELTA; 1.
KW Hydrolase; ATP synthesis; CF(1); Chloroplast; Hydrogen ion transport.
SQ SEQUENCE 183 AA; 20601 MW; 51C04FD66218AEB8 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLK 6
Db |||||
43 KSADLK 48

RESULT 9
ADH2_KLULA STANDARD; PRT; 348 AA.
ID ADH2_KLULA
AC P49383;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alcohol dehydrogenase II (EC 1.1.1.1).
GN ADH2.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
RX MEDLINE=92269769; PubMed=1588917;
RA Shain D.H., Salvadore C., Denis C.L.;
RT "Evolution of the alcohol dehydrogenase (ADH) genes in yeast:
RT characterization of a fourth ADH in Kluyveromyces lactis.";
RL Mol. Gen. Genet. 232:479-488(1992).
CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +
CC NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; X64397; CAA45739.1; -.
DR PIR; S20911; S20911.
DR InterPro; IPR002328; ADH_zinc.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
FT METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 67 67 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 98 98 ZINC 2 (BY SIMILARITY).
FT METAL 101 101 ZINC 2 (BY SIMILARITY).
FT METAL 104 104 ZINC 2 (BY SIMILARITY).
FT METAL 112 112 ZINC 2 (BY SIMILARITY).
FT METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).


```
SQ SEQUENCE 348 AA; 37097 MW; F3B64AE1F520689C CRC64;
Query Match 84.8%; Score 28; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLK 6
Db 164 KSADLK 169

RESULT 10
Y481 TREPA
ID Y481 TREPA STANDARD; PRT; 477 AA.
AC O83494;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0481.
GN TP0481.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OC NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388 (1998).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62333; CAA44207.1; -.
DR PIR; S35908; S35908.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003599; AAA_ATPase_cent.
DR InterPro; IPR001270; Chaprinin_clpA/B.
DR Pfam; PF00004; AAA; 1.
DR PRINTS; PR00300; CLPPROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPAB_1; FALSE NEG.
DR PROSITE; PS00871; CLPAB_2; FALSE NEG.
KW Chaperone; ATP-binding; Repeat; Plasmid.
FT DOMAIN 136 386
FT DOMAIN 437 629
FT NP_BIND 181 188 ATP (POTENTIAL).
FT NP_BIND 511 518 ATP (POTENTIAL).
SQ SEQUENCE 763 AA; 86018 MW; E92CCF2F629D9C14 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 763;
Best Local Similarity 71.4%; Pred. No. 92;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 411 KAADIKR 417

RESULT 12
K852 HUMAN
ID K852 HUMAN STANDARD; PRT; 970 AA.
AC Q9Y6X9; Q9JUF28; Q9Y6V2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein KIAA0852.
GN KIAA0852.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 5:355-364 (1998).
RN [2]
RP SEQUENCE FROM N.A.
```

```
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CNRZ 270;
RX MEDLINE=93247497; PubMed=8387149;
RA Huang D.C., Huang X.F., Novel G., Novel M.;
RT "Two genes present on a transposon-like structure in Lactococcus
RT lactis are involved in a clp-family proteolytic activity."
RL Mol. Microbiol. 7:957-965 (1993).
CC -!- FUNCTION: COULD BE THE ATP-DEPENDENT SPECIFICITY COMPONENT OF AN
CC ATP-DEPENDENT PROTEASE.
CC -!- SIMILARITY: Belongs to the clpA/clpB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62333; CAA44207.1; -.
DR PIR; S35908; S35908.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003599; AAA_ATPase_cent.
DR InterPro; IPR001270; Chaprinin_clpA/B.
DR Pfam; PF00004; AAA; 1.
DR PRINTS; PR00300; CLPPROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPAB_1; FALSE NEG.
DR PROSITE; PS00871; CLPAB_2; FALSE NEG.
KW Chaperone; ATP-binding; Repeat; Plasmid.
FT DOMAIN 136 386
FT DOMAIN 437 629
FT NP_BIND 181 188 ATP (POTENTIAL).
FT NP_BIND 511 518 ATP (POTENTIAL).
SQ SEQUENCE 763 AA; 86018 MW; E92CCF2F629D9C14 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 763;
Best Local Similarity 71.4%; Pred. No. 92;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
Db 411 KAADIKR 417

RESULT 12
K852 HUMAN
ID K852 HUMAN STANDARD; PRT; 970 AA.
AC Q9Y6X9; Q9JUF28; Q9Y6V2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein KIAA0852.
GN KIAA0852.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 5:355-364 (1998).
RN [2]
RP SEQUENCE FROM N.A.
```



```
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobby V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dharni P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 650-970 FROM N.A.
RC TISSUE=Testis;
RA Ottenwaelder B., Obermaier B., Mewes H.-W., Gassenhuber J.,
RA Wiemann S.;
```

```

Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
-!- TISSUE SPECIFICITY: Highly expressed in smooth muscle, pancreas
and testis.
-!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
gene model prediction.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB020659; BAA74875.1; -
EMBL; AC004542; AAC12954.1; ALT_SEQ.
EMBL; BC019257; AAH19257.1; -
EMBL; AL133637; CAB63760.1; -
PIR; T43455; T43455.
Coiled coil.
DOMAIN 220 300 COILED COIL (POTENTIAL).
DOMAIN 485 522 COILED COIL (POTENTIAL).
DOMAIN 679 699 COILED COIL (POTENTIAL).
DOMAIN 904 954 COILED COIL (POTENTIAL).
SEQUENCE 970 AA; 110724 MW; 459161807B8B53F6 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 970;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SADLKR 7
Db 723 SADLKR 728

RESULT 13
YHCR_BACSU STANDARD; PRT; 1217 AA.
AC P54602;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein yhcr precursor.
GN YHCR OR BSU09190.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124185; PubMed=8969498;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RT "A 22 kb DNA sequence in the cspB-glpPKD region at 75 degrees on the
Bacillus subtilis chromosome.";
RL Microbiology 142:3021-3026(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
```

RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan B., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE THERMONUCLEASE
CC FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE 5'-
CC NUCLEOTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X96983; CAA65702.1; -.
DR EMBL; Z99108; CAB12747.1; -.
DR PIR; F69823; F69823.
DR HSSP; P00644; 1SYB.
DR Subtilist; BG11596; yhcR.
DR InterPro; IPR008334; 5'-Nucleotidase C.
DR InterPro; IPR006146; 5'-Nucleotidase N.
DR InterPro; IPR006179; 5 nucleotidase.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR004843; M-peptidase.
DR InterPro; IPR006021; SNase.
DR InterPro; IPR004365; tRNA anti.
DR Pfam; PF02872; 5 nucleotidaseC; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF00149; Metallophos; 1.
DR Pfam; PF00565; SNase; 1.
DR Pfam; PF01336; tRNA anti; 2.
DR PRINTS; PR01607; APYRASEFAMILY.
DR SMART; SM00318; SNC; 1.
DR PROSITE; PS00785; 5_NUCLEOTIDASE_1; FALSE_NEG.
DR PROSITE; PS00786; 5_NUCLEOTIDASE_2; 1.
DR PROSITE; PS50830; TNASE_3; 1.
KW Hypothetical protein; Hydrolase; Signal; Complete proteome.
FT SIGNAL 1 46 POTENTIAL.
FT CHAIN 47 1217 HYPOTHETICAL PROTEIN YHCR.
FT DOMAIN 376 517 TNASE-LIKE.
SQ SEQUENCE 1217 AA; 132686 MW; 7973160CC898DC0B CRC64;

Query Match 84.8%; Score 28; DB 1; Length 1217;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLK 6
Db 358 KSADLK 363

RESULT 14
SOS2 MOUSE STANDARD; PRT; 1297 AA.
AC Q02384;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Son of sevenless protein homolog 2 (SOS-2) (msOS-2) (Fragment).
GN SOS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Eye;
RX MEDLINE=92335328; PubMed=1631150;
RA Bowtell D.D., Fu P., Simon M.A., Senior P.V.,
RT "Identification of murine homologues of the Drosophila son of
RT sevenless gene: potential activators of ras";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6511-6515(1992).
CC -!- FUNCTION: Promotes the exchange of Ras-bound GDP by GTP (By
CC similarity).
CC -!- TISSUE SPECIFICITY: Expressed in most embryonic and adult tissues.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 N-terminal Ras-GEF domain.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 Ras-GEF domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; Z11664; CAA77732.1; -.
DR PIR; S25714; S25714.
DR HSSP; Q62245; 1PMS.
DR MGD; MGI:98355; Sos2.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR001849; PH.
DR InterPro; IPR008937; Ras_GEF.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGRF_CDC25.
DR InterPro; IPR000219; RhogEF.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.
DR Pfam; PF00621; RhogEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 1.
DR SMART; SM00325; RhogEF; 1.
DR PROSITE; PS00741; DH_1; FALSE_NEG.
DR PROSITE; PS0010; DH_2; 1.
DR PROSITE; PS00720; RASGEF; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50009; RASGEF_CAT; 1.
DR PROSITE; PS50212; RASGEF_NTER; 1.
KW Guanine-nucleotide releasing factor.
FT NON_TER 1 1
FT DOMAIN 164 353 DH.
FT DOMAIN 407 510 PH.
FT DOMAIN 560 704 N-TERMINAL RAS-GEF.
FT DOMAIN 743 982 RAS-GEF.
FT DOMAIN 1145 1148 POLY-PRO.
FT DOMAIN 1164 1173 POLY-PRO.
SQ SEQUENCE 1297 AA; 148846 MW; 3A53F8515BF9D625 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 1297;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SADLXR 7
Db 634 SADLXR 639

```
RESULT 15
SOS2_HUMAN STANDARD; PRT; 1332 AA.
ID SOS2_HUMAN
AC Q07890; Q15503;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Son of sevenless protein homolog 2 (SOS-2).
GN SOS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93262494; PubMed=8493579;
RA Chardin P., Camonis J.H., Gale N.W., van Aelst L., Wigler M.H.,
RA Bar-Sagi D.;
RT "Human Sos1: a guanine nucleotide exchange factor for Ras that binds
RT to GRB2."
RL Science 260:1338-1343 (1993).
RN [2]
RP SEQUENCE OF 724-1296 FROM N.A.
RC TISSUE=Placenta;
RA Fath I., Apio F., Dutrillaux B., Tocque B.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Promotes the exchange of Ras-bound GDP by GTP (By
CC similarity).
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 N-terminal Ras-GEF domain.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 Ras-GEF domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L13858; AAA35914.1; -.
CC EMBL; L20686; AAA91852.1; -.
CC HSSP; Q62245; 1PMS.
CC Genew; HGNC:11188; SOS2.
CC MIM; 601247; -.
CC GO; GO:0005085; P:guanyl-nucleotide exchange factor activity; NAS.
CC GO; GO:0007264; P:small GTPase mediated signal transduction; NAS.
CC InterPro; IPR001331; GDS_CDC24.
CC InterPro; IPR001849; PH.
CC InterPro; IPR008937; Ras_GEF.
CC InterPro; IPR000651; RasGEFN.
CC InterPro; IPR001895; RasGRF_CDC25.
CC InterPro; IPR000219; RhogEF.
CC Pfam; PF00169; PH; 1.
CC Pfam; PF00617; RasGEF; 1.
CC Pfam; PF00618; RasGEFN; 1.
CC Pfam; PF00621; RhogEF; 1.
CC SMART; SM00233; PH; 1.
CC SMART; SM00147; RasGEF; 1.
CC SMART; SM00229; RasGEFN; 1.
CC SMART; SM00325; RhogEF; 1.
CC PROSITE; PS00741; DH_1; FALSE_NEG.
CC PROSITE; PS0010; DH_2; 1.
CC PROSITE; PS00720; RasGEF; 1.
CC PROSITE; PS50003; PH DOMAIN; 1.
CC PROSITE; PS50009; RasGEF_CAT; 1.
CC PROSITE; PS50212; RasGEF_NTER; 1.
CC Guanine-nucleotide releasing factor.
FT DOMAIN 198 388
FT 442 546
FT PH.
```

```
FT DOMAIN 595 739 N-TERMINAL RAS-GEF.
FT DOMAIN 778 1017 RAS-GEF.
FT DOMAIN 755 758 POLY-PRO.
FT DOMAIN 1180 1183 POLY-PRO.
FT DOMAIN 1203 1208 POLY-PRO.
FT CONFLICT 778 778 D -> H (IN REF. 2).
FT CONFLICT 861 861 Q -> R (IN REF. 2).
FT CONFLICT 948 948 R -> K (IN REF. 2).
FT CONFLICT 999 999 S -> C (IN REF. 2).
FT CONFLICT 1032 1032 S -> E (IN REF. 2).
FT CONFLICT 1042 1042 T -> A (IN REF. 2).
FT CONFLICT 1112 1114 CGS -> LC (IN REF. 2).
FT CONFLICT 1159 1159 A -> G (IN REF. 2).
FT CONFLICT 1296 1296 Q -> E (IN REF. 2).
SQ SEQUENCE 1332 AA; 153030 MW; EE4BA9E52FA85AC9 CRC64;
```

Query Match 84.8%; Score 28; DB 1; Length 1332;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SADLKR 7
Db 669 SADLKR 674

Search completed: May 24, 2004, 14:51:22
Job time : 13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 14:47:48 ; Search time 39 Seconds
(without alignments)
56.631 Million cell updates/sec

Title: US-09-807-665-41
Perfect score: 33
Sequence: 1 KSADLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL 25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	90.9	160	16 Q88SX2	Q88sx2 lactobacill
2	30	90.9	328	3 Q9UR34	Q9ur34 schizosacch
3	30	90.9	1626	16 Q9RY7	Q9ry7 deinococcus
4	30	90.9	2100	12 Q9DWE2	Q9dwe2 rat cytomeg
5	30	90.9	2403	5 Q8MLD5	Q8mld5 drosophila
6	30	90.9	4463	5 Q8MLD8	Q8mld8 drosophila
7	30	90.9	7210	5 Q9V7G8	Q9v7g8 drosophila
8	30	90.9	9270	5 Q8MLD9	Q8mld9 drosophila
9	29	87.9	139	16 Q832V5	Q832v5 enterococcu
10	29	87.9	186	16 Q89X71	Q89x71 bradyrhizob
11	29	87.9	333	16 Q8RD61	Q8rd61 thermoanaer
12	29	87.9	373	16 Q821Z1	Q821z1 chlamydomon
13	29	87.9	454	10 Q8H6A4	Q8h6a4 stevia reba
14	29	87.9	648	5 Q76572	Q76572 caenorhabdi
15	29	87.9	720	10 Q7XKX9	Q7xkx9 oryza sativ
16	29	87.9	1029	16 Q9A9L0	Q9a9l0 caulobacter

Q965j0 caenorhabdi
Q9ewb8 streptomyce
Q7vka3 haemophilus
Q9hky4 thermoplasm
Q854z0 mycobacteri
Q813i6 bacillus ce
Q8nuu0 staphylococ
Q8cn28 staphylococ
Q88zr2 lactobacill
Q7wx27 alcaligenes
Q89g55 bradyrhizob
Q8gni4 bradyrhizob
Q82834 pseudomonas
Q8svj6 encephalito
Q9zxh3 oenococcus
Q9lt79 arabidopsis
Q8puj3 methanosarc
Q8rbx3 thermoanaer
Q8rbg4 thermoanaer
Q8f7s5 leptospira
Q93dt0 streptococc
Q7wsn2 streptococc
Q8e777 streptococc
Q8e0k4 streptococc
Q9lpy5 arabidopsis
Q7uv96 rhodopirell
Q8sav5 oryza sativ
Q74964 schizosacch
Q85942 sphingomona

ALIGNMENTS

RESULT 1

Q88SX2 ID Q88SX2 PRELIMINARY; PRT; 160 AA.
AC Q88SX2;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Ferric uptake regulator.
GN FUR OR LP 3247.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RA MEDLINE=22480296; PubMed=12565566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935261; CAD65387.1; -.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR002481; FUR.
DR Pfam; PF01475; FUR; 1.
KW Complete proteome.
SQ SEQUENCE 160 AA; 18130 MW; F37A76FD356A5819 CRC64;

Query Match 90.9%; Score 30; DB 16; Length 160;
Best Local Similarity 85.7%; Pred. No. 62;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7

DB 147 KTADLKR 153

Query Match 90.9%; Score 30; DB 16; Length 1626;
Best Local Similarity 85.7%; Pred. No. 6.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 522 KSADLRR 528

RESULT 4

Q9DWE2
ID Q9DWE2 PRELIMINARY; PRT; 2100 AA.
AC Q9DWE2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE PR48.
GN R48.

OS Rat cytomegalovirus (strain Maastricht).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Muromegalovirus.
OX NCBI_TaxID=79700;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Maastricht;
RX MEDLINE=20366325; PubMed=10906222;
RA Vink C., Beuken E., Bruggeman C.A.;
RT "Complete DNA sequence of the rat cytomegalovirus genome.";
RL J. Virol. 74:7656-7665(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Maastricht;
RX MEDLINE=20473137; PubMed=11018281;
RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A.;
RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a
RT spliced transcript.";
RL Virus Res. 69:119-130(2000).
DR EMBL; AF232689; AAF99147.1; -;
DR InterPro; IPR006928; Herpes_teg_N.
DR Pfam; PF04843; Herpes_teg_N; 1.
SQ SEQUENCE 2100 AA; 234495 MW; C242AAAB526877BB CRC64;

Query Match 90.9%; Score 30; DB 12; Length 2100;
Best Local Similarity 85.7%; Pred. No. 8.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 933 RSADLKR 939

RESULT 5

Q8MLD5
ID Q8MLD5 PRELIMINARY; PRT; 2403 AA.
AC Q8MLD5;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE CG18255-PE.
GN STRN-MLCK OR CG8304 OR CG18255.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RESULT 2

Q9UR34
ID Q9UR34 PRELIMINARY; PRT; 328 AA.
AC Q9UR34;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE SUR4 family protein.
GN SPAC1639.01C OR SPAC806.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Wedler H., Duesterhoeft A., McDougall R.C., Rajandream M.A.,
RA Barrell B.G.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117213; CAB55289.1; -;
DR EMBL; AL117212; CAB55288.1; -;
DR PIR; T37734; T37734.
DR GeneDB SPombe; SPAC1639.01C; -;
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR02076; GNS1_SUR4.
DR Pfam; PF01151; ELO; 1.
SQ SEQUENCE 328 AA; 37577 MW; 36443BC29CFB9D4B CRC64;

Query Match 90.9%; Score 30; DB 3; Length 328;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 291 KNADLKR 297

RESULT 3

Q9RYV7
ID Q9RYV7 PRELIMINARY; PRT; 1626 AA.
AC Q9RYV7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein DRA0166.
GN DRA0166.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001862; AAF12309.1; -;
DR PIR; A75613; A75613.
DR TIGR; DRA0166; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1626 AA; 179055 MW; CCAEB677AB8903AF CRC64;

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston X.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Bazzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 15 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 DR EMBL; AE003808; AAM70939.1; -
 DR FlyBase; FBgn0013988; Strn-Mlck.

GO; GO:0004685; F:calcium/calmodulin-dependent protein kinase. .; IDA.
GO; GO:0004687; F:myosin-light-chain kinase activity; IDA.
InterPro; IPR007110; Ig-like.
InterPro; IPR003598; Ig_c2.
InterPro; IPR003006; Ig_MHC.
Pfam; PF00047; Ig; 14.
SMART; SM00408; IGC2; 16.
DR PROSITE; PS0835; IG LIKE; 14.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 2403 AA; 269233 MW; 60181230959FF9F4 CRC64;

Query Match 90.9%; Score 30; DB 5; Length 2403;
Best Local Similarity 85.7%; Pred. No. 1e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
|:|||||
Db 1931 KAADLKR 1937

RESULT 6

Q8MLD8 PRELIMINARY; PRT; 4463 AA.

AC Q8MLD8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG18255-PC.
GN STRN-MLCK OR CG8304 OR CG18255.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodis A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Noshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpkins M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaninker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003808; AAM70937.1; -;
DR FlyBase; FBgn0013988; Strn-Mlck.
DR GO; GO:0004685; F:calcium/calmodulin-dependent protein kinase. . . ; IDA.
DR GO; GO:0004687; F:myosin-light-chain kinase activity; IDA.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR InterPro; IPR007119; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; ig; 23.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGC2; 25.
DR SMART; SM00406; IGV; 3.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyTKc; 1.
DR PROSITE; PS50835; IG LIKE; 22.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Immunoglobulin domain; Transferase.
SQ SEQUENCE 4463 AA; 497000 MW; DE2E691FCE94B44B CRC64;

Query Match 90.9%; Score 30; DB 5; Length 4463;
Best Local Similarity 85.7%; Pred. No. 1.9e+03;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSADLKR 7
Db 1931 KAADLKR 1937
RESULT 7
Q9V7G8 PRELIMINARY; PRT; 7210 AA.
ID Q9V7G8
AC Q9V7G8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG18255 protein.
GN STRN-MLCK OR CG8304 OR CG18255.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003808; F:calcium/calmodulin-dependent protein kinase. . ; IDA.
DR GO; GO:0004685; F:calcium/calmodulin-dependent protein kinase activity; IDA.
DR GO; GO:0004687; F:myosin-light-chain kinase activity; IDA.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; ig; 19.
DR SMART; SM00409; IG; 21.
DR SMART; SM00408; IGC2; 20.
DR PROSITE; PS00835; IG LIKE; 19.
DR PROSITE; PS00290; IG_MHC; 1.
DR KW Immunoglobulin domain.
SQ SEQUENCE 7210 AA; 808965 MW; 6B2E7395C07140D2 CRC64;

Query Match 90.9%; Score 30; DB 5; Length 7210;
Best Local Similarity 85.7%; Pred. No. 3.1e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7
|:|||||
Db 6738 KAADLKR 6744

RESULT 8
Q8MLD9 PRELIMINARY; PRT; 9270 AA.
AC Q8MLD9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG18255-PA.
GN STRN-MLCK OR CG8304 OR CG18255.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Brill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dudson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003808; F:calcium/calmodulin-dependent protein kinase. . ; IDA.
DR GO; GO:0004685; F:calcium/calmodulin-dependent protein kinase activity; IDA.
DR GO; GO:0004687; F:myosin-light-chain kinase activity; IDA.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; ig; 19.
DR SMART; SM00409; IG; 21.
DR SMART; SM00408; IGC2; 20.
DR PROSITE; PS00835; IG LIKE; 19.
DR PROSITE; PS00290; IG_MHC; 1.
DR KW Immunoglobulin domain.
SQ SEQUENCE 7210 AA; 808965 MW; 6B2E7395C07140D2 CRC64;

DR GO; GO:0004687; F:myosin-light-chain kinase activity; IDA.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; ig; 28.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00408; IGC2; 29.
DR SMART; SM00406; IGV; 4.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00835; IG_LIKE; 27.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Immunoglobulin domain; Transferase.
SQ SEQUENCE 9270 AA; 1036749 MW; 0DD82D040B27C2AD CRC64;

Query Match 90.9%; Score 30; DB 5; Length 9270;
Best Local Similarity 85.7%; Pred. No. 4.1e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 6738 KAADLKR 6744

RESULT 9
Q832V5
ID Q832V5 PRELIMINARY; PRT; 139 AA.
AC Q832V5;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN EF2113.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Gill S.R., Heidelberg J.F.,
RA Read T.D., Fouts D.E., Eisen J.A., Brinkac L., Beanan M.,
RA Tettelin H., Dodson R.J., Umayam L., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
Enterococcus faecalis."
RL Science 299:2071-2074(2003).
DR EMBL; AE016953; AAO81846.1; -.
DR TIGR; EF2113; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 139 AA; 15431 MW; F757A4EC2A159D9C CRC64;

Query Match 87.9%; Score 29; DB 16; Length 139;
Best Local Similarity 85.7%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 36 KAADLKR 42

RESULT 10
Q89X71
ID Q89X71 PRELIMINARY; PRT; 186 AA.
AC Q89X71;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE ATP synthase delta chain.
GN ATP OR BL0443.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005936; BAC45708.1; -.
DR GO; GO:0006754; P:ATP biosynthesis; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR000711; ATPsynt_OSCP.
DR Pfam; PF00213; OSCP; 1.
DR PRINTS; PR00125; ATPASEDELTA.
DR TIGRFAMS; TIGR01145; ATP_synt_delta; 1.
DR PROSITE; PS00389; ATPASE_DELTA; 1.
KW Complete proteome.
SQ SEQUENCE 186 AA; 19670 MW; 4125BD0CD9D60750 CRC64;

Query Match 87.9%; Score 29; DB 16; Length 186;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSADLKR 7
Db 45 ESADLKR 51

RESULT 11
Q8RD61
ID Q8RD61 PRELIMINARY; PRT; 333 AA.
AC Q8RD61;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Thiamine pyrophosphate-dependent dehydrogenases, E1 component alpha
subunit.
GN ACOA OR TTE0186.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
DR EMBL; AE012992; AAM23487.1; -.
DR GO; GO:0016624; F:oxidoreductase activity, acting on the alde. . .; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001017; Dehydrogenase_E1.

DR Pfam; PF00676; El_dehydrog; 1.
KW Complete proteome.
SQ SEQUENCE 333 AA; 37059 MW; 1AB8C53321E7C135 CRC64;

Query Match 87.9%; Score 29; DB 16; Length 333;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 84 KGADLKR 90

RESULT 12

Q821Z1
ID Q821Z1 PRELIMINARY; PRT; 373 AA.

AC Q821Z1;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN CCA00794.
OS Chlamydomophila caviae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12682364;
RA Read T.D.; Myers G.S.A.; Brunham R.C.; Nelson W.C.; Paulsen I.T.;
RA Heidelberg J.; Holtzapple E.; Khouri H.; Federova N.B.; Carty H.A.;
RA Umayam L.A.; Haft D.H.; Peterson J.; Beanan M.J.; White O.;
RA Salzberg S.L.; Hsia R.-C.; McClarty G.; Rank R.G.; Bavoil P.M.;
RA Fraser C.M.;
RT "Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae.";
RL Nucleic Acids Res. 31:2134-2147(2003).
DR EMBL; AE016997; AAP05535.1; -.
DR TIGR; CCA00794; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 373 AA; 41724 MW; E070E6171DA8E6C CRC64;

Query Match 87.9%; Score 29; DB 16; Length 373;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 191 ESADLKR 197

RESULT 13

Q8H6A4
ID Q8H6A4 PRELIMINARY; PRT; 454 AA.

AC Q8H6A4;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE UDP-glucosyltransferase.
OS Stevia rebaudiana (Stevia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC campanulids; Asterales; Asteraceae; Eupatorieae; Stevia.
OX NCBI_TaxID=55670;
RN [1]
RP SEQUENCE FROM N.A.
RA Ma L.; Chen M.; Chen L.;
RT "Molecular cloning and characterization of Stevia rebaudiana UDP-
RT glucosyltransferase.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF548025; AAN40684.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.

KW Transferase.
SQ SEQUENCE 454 AA; 50826 MW; FOCA66D07A458E5A CRC64;

Query Match 87.9%; Score 29; DB 10; Length 454;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 304 KSAELKR 310

RESULT 14

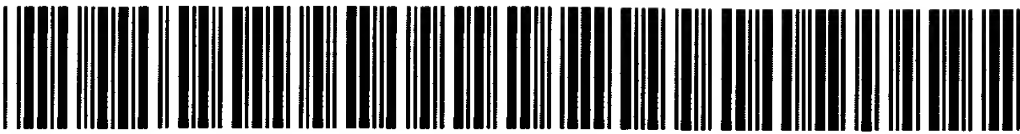
O76572
ID O76572 PRELIMINARY; PRT; 648 AA.

AC O76572;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN K07D4.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Henkhaus J.; Wohldmann P.;
RT "The sequence of C. elegans cosmid K07D4.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AF077534; AAC26284.1; -.
DR PIR; T33339; T33339.
DR WormPep; K07D4.7b; CE19531.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Hypothetical protein; SH3 domain.
SQ SEQUENCE 648 AA; 73698 MW; 65F7244CA13621DD CRC64;

Query Match 87.9%; Score 29; DB 5; Length 648;
Best Local Similarity 85.7%; Pred. No. 4.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSADLKR 7

Db 380 KDADLKR 386



Creation date: 06-10-2004
Indexing Officer: VKHOMRASAVONG - VIVIAN KHOMRASAVONG
Team: OIPE Scanning
Dossier: 10240758

Legal Date: 06-09-2004

No.	Doccode	Number of pages
1	CTNF	6
2	892	1
3	1449	1
4	SRFW	1
5	FWCLM	1

Total number of pages: 10

Remarks:

Order of re-scan issued on